

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number  
**WO 01/18225 A1**

(51) International Patent Classification<sup>7</sup>: C12N 15/90, A01K 67/027

(21) International Application Number: PCT/US99/30078

(22) International Filing Date:  
16 December 1999 (16.12.1999)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/152,522 3 September 1999 (03.09.1999) US

(71) Applicant: XENOGEN CORPORATION [US/US]; 860 Atlantic Avenue, Alameda, CA 94501 (US).

(72) Inventor: ZHANG, Ning; Xenogen Corporation, 860 Atlantic Avenue, Alameda, CA 94501 (US).

(74) Agent: SHOLTZ, Charles, K.; Xenogen Corporation, 860 Atlantic Avenue, Alameda, CA 94501 (US).

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW.

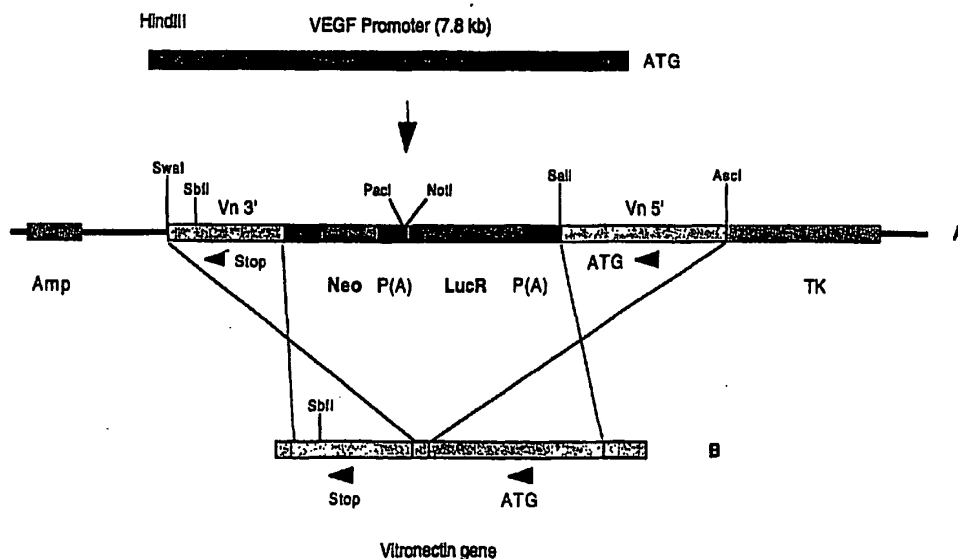
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— With international search report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TARGETING CONSTRUCTS AND TRANSGENIC ANIMALS PRODUCED THEREWITH



(57) Abstract: The present invention teaches targeting constructs and methods of use thereof for creating transgenic animals in which at least one single-copy, non-essential gene is replaced with a reporter expression cassette, for example, a luciferase gene operably linked to a promoter heterologous to the single-copy, non-essential gene. Thus, the present invention provides novel methods and vector constructs useful for the generation of transgenic animals. The invention further includes methods of using these animals.

WO 01/18225 A1

## TARGETING CONSTRUCTS AND TRANSGENIC ANIMALS PRODUCED THEREWITH

5

### TECHNICAL FIELD

This invention is in the field of molecular biology and medicine. More specifically, it relates to novel vector constructs and methods of use thereof for introducing heterologous polynucleotides into a host cell. Further, the invention relates to vector constructs and methods of use thereof to generate transgenic organisms, particularly transgenic mice.

### BACKGROUND

In recent years, mouse geneticists have succeeded in creating transgenic animals by manipulating the genes of developing embryos and introducing foreign genes into these embryos. Once these genes have integrated into the genome of the recipient embryo, the resulting embryos or adult animals can be analyzed to determine the function of the gene.

Traditionally, transgenic mice have been generated through DNA microinjection approach. Such an approach leads to the creation of "founder" mice having at least one copy of the transgene randomly integrated into the genome. Because neither the copy number nor the integration sites can be controlled, transgenic mice generated by this method are genetically different from each other. The expression of transgenes in such transgenic mice are not uniform because of, for example, the difference in copy numbers of the transgene. Furthermore, chromosomal location of the transgene often affects the expression level of the transgene. For most of *in vivo* studies, particularly ones where it is desirable to compare levels of gene expression across different animals, it is important to have the mice with little or no genetic variation (i.e., isogenic mice) in order to reduce the systematic error.

One way of controlling copy number and the chromosomal integration site is by using targeting constructs to create transgenic animals. U.S. Patent Nos. 5,464,764 and 5,487,992 describe this type of transgenic animal in which a gene of interest is deleted or mutated sufficiently to disrupt its function. These "knock-out" animals are made by taking advantage of the phenomena of homologous recombination. (See, also U.S.

Patent Nos. 5,631,153 and 5,627,059). Using the "knock-out" approach, when the vector is introduced into the embryonic stem cell, a sequence becomes integrated into a target gene in the genome via homologous recombination. The integration disrupts the function of the target gene allows for examination of the phenotype resulting from the disruption of the gene. Neither of the above methods, however, provide a predictable approach for the generation of gene expression reporter transgenic animals which can be used for quantitative comparisons of gene expression across different animals transfected with the same construct, or for quantitative comparisons of the relative levels of expression of two or more different genes.

10       The present invention solves this and other problems by providing transgenic animals in which at least one single-copy, non-essential gene is replaced with a reporter expression cassette, for example, a luciferase gene operably linked to a heterologous promoter. Thus, the present invention provides novel methods and vector constructs useful for the generation of transgenic animals. The invention further includes methods  
15 of using these animals.

#### SUMMARY OF THE INVENTION

The transgenic animals described herein are useful, for example, when studying *in vivo* regulation of selected genes. Also described herein are methods of generating  
20 populations of substantially isogenic transgenic animals, as well as, vectors useful in these methods.

Accordingly, in one embodiment, the subject invention is directed to a transgenic, non-human mammal, for example, a rodent such as a mouse. The mammal comprises at least one single-copy, non-essential gene in its genome, wherein (i) at least a portion of  
25 at least one single-copy, non-essential gene is replaced by polynucleotide sequences heterologous to the gene, and (ii) the polynucleotide sequences comprise a first expression cassette which has been introduced into the mammal or an ancestor of the mammal, at an embryonic stage. The first expression cassette typically comprises a first selectable marker, a first transcriptional promoter element heterologous to the gene, and  
30 light-generating protein coding sequences. The light-generating protein coding sequences are operably linked to the promoter element.

The single-copy, non-essential gene may be selected, for example, from the group consisting of vitronectin, *fosB*, and galactin 3 and the first selectable marker may be selected from the group consisting of neomycin phosphotransferase II, xanthine-guanine phosphoribosyltransferase, hygromycin-B-phosphotransferase, chloramphenicol  
5 acetyltransferase, and adeninephosphoribosyl transferase.

In alternative embodiments, the first transcriptional promoter element is an inducible promoter, a repressible promoter, or a constitutive promoter, and may be selected from the group consisting of VEGF, VEGFR, and TIE2.

In additional embodiments, the transgenic, non-human mammal described above  
10 comprises a second single-copy, non-essential gene in its genome, wherein (i) at least a portion of the second single-copy, non-essential gene is replaced by polynucleotide sequences heterologous to the second gene, and (ii) the polynucleotide sequences comprise a second expression cassette which has been introduced into the mammal or an ancestor of the mammal, at an embryonic stage. The second expression cassette  
15 typically comprises a second selectable marker, a second transcriptional promoter element heterologous to the second gene, and light generating protein coding sequences. The light generating protein coding sequences are operably linked to the promoter element.

The first and second transcriptional promoter elements and selectable markers  
20 may be the same or different and the light generating protein in the first expression cassette can produce a different color of light relative to the light generating protein in the second expression cassette.

In yet a further embodiment, the invention is directed to a method of producing a transgenic, non-human mammal, such as a mouse. The mammal has at least one single-  
25 copy, non-essential gene in its genome. The method comprises

transfecting an embryonic stem cell of the mammal with a linear vector comprising

(a) a first selectable marker and a reporter expression cassette, the reporter expression cassette comprising a transcriptional promoter element operably linked to a  
30 light generating protein coding sequence, and

(b) targeting polynucleotide sequences homologous to a single-copy, non-essential gene in said mammal's genome, the targeting polynucleotide sequences

flanking (a), wherein (i) the length of the polynucleotide sequences are sufficient to facilitate homologous recombination between the vector and the single-copy, non-essential gene, and (ii) the transcriptional promoter element is heterologous to the single-copy, non-essential gene;

5        selecting embryonic stem cells which each have the first selectable marker and reporter expression cassette integrated into its genome;

          injecting the embryonic stem cells into a host embryo,

          implanting the embryo in a foster mother,

          maintaining the foster mother under conditions which allow production of an  
10        offspring that is a transgenic, non-human mammal carrying the reporter expression cassette.

          In certain embodiments, the offspring is capable of germline transmission of the reporter expression cassette and the method may further comprise breeding the offspring with a mammal which is substantially isogenic with the embryonic stem cells, such that  
15        the breeding yields transgenic F1 offspring carrying the reporter cassette. In particular embodiments, the method comprises breeding the first F1 offspring carrying the reporter cassette with a second F1 offspring carrying the reporter cassette, wherein the breeding yields transgenic F2 offspring carrying the reporter cassette.

          In additional embodiments, the embryonic stems cells may be derived from a  
20        mouse having a dark coat color, the mammal substantially isogenic with the embryonic stem cells may have a light coat color, and/or the F2 offspring carrying the reporter cassette may have a light coat color. In particular embodiments, the embryonic stems cells are derived from a C57BL/6 mouse having a dark coat color, and the mammal substantially isogenic with the embryonic stem cells is a C57BL/6-Tyr C2j/+ mouse  
25        having a light coat color.

          In still a further embodiment, the subject invention is directed to a vector for use in generating a transgenic non-human mammal, for example, a rodent such as a mouse. The mammal has at least one single-copy, non-essential gene in its genome. The vector comprises

30        (a) a first selectable marker and a reporter expression cassette, the reporter expression cassette comprising a transcriptional promoter element operably linked to a light generating protein coding sequence, and

(b) targeting polynucleotide sequences homologous to a single-copy, non-essential gene in the mammal's genome, the targeting polynucleotide sequences flanking (a), wherein (i) the length of the targeting polynucleotide sequences are sufficient to facilitate homologous recombination between the vector and the single-copy, non-essential gene, and (ii) the transcriptional promoter element is heterologous to the single-copy, non-essential gene.

In certain embodiments, the first selectable marker provides a positive selection and may be selected from the group consisting of neomycin phosphotransferase II, xanthine-guanine phosphoribosyltransferase, hygromycin-B-phosphotransferase, chloramphenicol acetyltransferase, and adenine phosphoribosyl transferase. Additionally, the transcriptional promoter element may be an inducible promoter, a repressible promoter, or a constitutive promoter, and may be selected from the group consisting of VEGF, VEGFR, and TIE2.

In alternative embodiments, the vector further comprises a second selectable marker and at least one target polynucleotide sequence is located between the second selectable marker and the first selectable marker. Additionally, the second selectable marker may provide a negative selection and may be selected from the group consisting of adenosine deaminase, thymidine kinase, and dihydrofolate reductase.

The vectors described above may be circular and may contain at least one restriction site whose cleavage results in a linear vector having the following arrangement of elements: target polynucleotide sequence - (a) - targeting polynucleotide sequences or target polynucleotide sequence - (a) - targeting polynucleotide sequences - (second selectable marker).

The coding sequences of the reporter expression cassette present in the vector may comprise codons that are optimal for expression in a host system into which the expression cassette is to be introduced. Additionally, the targeting polynucleotide sequences from single-copy, non-essential genes may be selected from the group consisting of vitronectin, *fosB*, and galactin 3.

The light-generating protein in the mammals and methods described above may be derived from either procaryotic or eucaryotic sources and, in particularly preferred embodiments, the light generating protein is a luciferase.

These and other embodiments of the present invention will be apparent to those of skill in the art in view of the teachings herein.

### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic depicting construction of the pTK53 vector.

Polynucleotides encoding PGK-P, Neo and TK and 5' and 3' linkers are introduced into  
5 a pKS backbone to produce the vector designated pTK53.

Figure 2 is schematic depicting construction of the pTK-LucR and pTK-LucYG vectors. For pTK-LucR, a polynucleotide encoding LucR is introduced into pTK53.

Thus, the pTK-LucR construct contains the PGK-P gene, a neomycin (Neo<sup>r</sup>) gene, a thymidine kinase (TK) gene and sequence encoding red luciferase (Luc-R). For pTK-  
10 LucYG, a polynucleotide encoding LucYG is introduced into pTK53. Thus, the pTK-LucYG construct contains the PGK-P gene, a neomycin (Neo<sup>r</sup>) gene, a thymidine kinase (TK) gene and a sequence encoding yellow-green luciferase (Luc-YG).

Figures 3A is a schematic depicting the vector pTKLR-Vn. Sequences homologous to the vitronectin gene are inserted into pTK-LucR such that they flank the  
15 Neo<sup>r</sup> gene and the Luc-R coding sequence. Figure 3B is a schematic depicting targeting of the linearized pTKLR-Vn vector to the vitronectin chromosomal locus. The VEGF promoter is cloned into the polylinkers between Neo and Luc-R. Upon homologous recombination, the Neo-VEGF-LucR transgene is inserted into the Vn gene. In the figure, (A) shows the targeting vector pTKLR-Vn and (B) shows the mouse vitronectin  
20 gene. In the figure, Neo – neomycin resistance encoding sequences; TK – thymidine kinase encoding sequences; LucR – red luciferase from pGL3Red (Dr. Christopher Contag, Stanford University, Stanford, CA). Regions bearing Vn gene translational start and stop codons are indicated with arrows. Poly(A) sequences are placed upstream of the polylinker to prevent or minimize read-through translation. Figure 3C shows the  
25 nucleotide sequence of vitronectin (SEQ ID NO:38).

Figure 4A is a schematic depicting the vector pTKLG-Fos. Sequences homologous to the FosB gene are inserted into pTK-LucYG such that they flank the Neo<sup>r</sup> gene and the Luc-YG coding sequence. Figure 4B shows the nucleotide sequence of FosB (SEQ ID NO:39).

30 Figure 5A is a schematic depicting targeting of the linearized pTKLG-Fos vector to the FosB chromosomal locus. The VEGFR2 promoter is cloned into the polylinkers between Neo and Luc-YG. Upon homologous recombination, the Neo-VEGFR2-LucYG



transgene will be inserted into a sequence associated with production of FosB. In the figure, (A) shows the targeting vector, and (B) shows the mouse target gene. In the figure, Neo – neomycin resistance encoding sequences; TK – thymidine kinase encoding sequences; LucYG – yellow green luciferase from pGL3-control vector (Promega, Madison, WI). Regions bearing FosB gene translational start and stop codons are indicated with arrows. Poly(A) sequences are placed upstream of the polylinker to prevent or minimize read-through translation. Figure 5B is a schematic depicting targeting of the linearized pTKLG-Fos vector to the FosB chromosomal locus. The TIE2 promoter is cloned into the polylinkers between Neo and Luc-R. Upon homologous recombination, the Neo-Tie2-LucYG transgene is inserted into the FosB gene. In the figure, (A) shows the targeting vector, and (B) shows the mouse target gene. In the figure, Neo – neomycin resistance encoding sequences; TK – thymidine kinase encoding sequences; LucYG – yellow green luciferase from pGL3-control vector (Promega). Regions bearing FosB gene translational start and stop codons are indicated with arrows. Poly(A) sequences are placed upstream of the polylinker to prevent or minimize read-through translation.

Figure 6 depicts PCR conditions for genomic screening for promoters useful in exemplary targeting constructs of the present invention.

Figure 7 depicts generation of targeted transgenic mice using the targeting vectors described herein.

Figure 8 depicts of schematic representation of Southern blot analysis of homologous DNA recombination between pTKLG-Fos targeting vector and the FosB gene.

Figure 9 depicts generation of targeted transgenic mice, using the targeting vectors described herein, and crosses using such transgenics as well as their offspring (F1, first generation; F2, second generation).

Figure 10 depicts crosses using transgenic mice of the present invention to generate dual luciferase transgenic mice.

### MODES FOR CARRYING OUT THE INVENTION

Throughout this application, various publications, patents, and published patent applications are referred to by an identifying citation. The disclosures of these publications, patents, and published patent specifications referenced in this application are included to more fully describe the state of the art to which this invention pertains.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, cell biology and recombinant DNA, which are within the skill of the art. *See, e.g.*, Sambrook, Fritsch, and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, (F.M. Ausubel et al. eds., 1987); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (M.J. McPherson, B.D. Hames and G.R. Taylor eds., 1995) and ANIMAL CELL CULTURE (R.I. Freshney. Ed., 1987).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

#### **Definitions**

As used herein, certain terms will have specific meanings.

The terms "nucleic acid molecule" and "polynucleotide" are used interchangeably to and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown.

Non-limiting examples of polynucleotides include a gene, a gene fragment, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers.

A polynucleotide is typically composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); and thymine (T) (uracil (U) for thymine (T) when the polynucleotide is RNA). Thus, the term polynucleotide sequence is the alphabetical representation of a polynucleotide molecule. This alphabetical

representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a  
5 nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide, for example, *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are typically determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but  
10 is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence. Other "control elements" may also be associated with a coding sequence. A DNA sequence encoding a polypeptide can be optimized for expression in a selected cell by using the  
15 codons preferred by the selected cell to represent the DNA copy of the desired polypeptide coding sequence. "Encoded by" refers to a nucleic acid sequence which codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even more preferably at least 15 to 20 amino acids from a  
20 polypeptide encoded by the nucleic acid sequence. Also encompassed are polypeptide sequences which are immunologically identifiable with a polypeptide encoded by the sequence.

Typical "control elements", include, but are not limited to, transcription promoters, transcription enhancer elements, transcription termination signals,  
25 polyadenylation sequences (located 3' to the translation stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), translation enhancing sequences, and translation termination sequences. Transcription promoters can include inducible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein,  
30 etc.), repressible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), and constitutive promoters.

"Expression enhancing sequences" typically refer to control elements that improve transcription or translation of a polynucleotide relative to the expression level in the absence of such control elements (for example, promoters, promoter enhancers, enhancer elements, and translational enhancers (e.g., Shine and Delagarno sequences)).

5 "Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof which is essentially free, e.g., contains less than about 50%, preferably less than about 70%, and more preferably less than about 90%, of the protein with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides of interest are well-known in the art and include, for example, disruption of the cell  
10 containing the polynucleotide with a chaotropic agent and separation of the polynucleotide(s) and proteins by ion-exchange chromatography, affinity chromatography and sedimentation according to density.

A "heterologous sequence" as used herein is typically refers to either (i) a nucleic acid sequence that is not normally found in the cell or organism of interest, or (ii) a  
15 nucleic acid sequence introduced at a genomic site wherein the nucleic acid sequence does not normally occur in nature at that site. For example, a DNA sequence encoding a polypeptide can be obtained from yeast and introduced into a bacterial cell. In this case the yeast DNA sequence is "heterologous" to the native DNA of the bacterial cell. Alternatively, a promoter sequence from a Tie2 gene can be introduced into the genomic  
20 location of a *fosB* gene. In this case the Tie2 promoter sequence is "heterologous" to the native *fosB* genomic sequence.

A "polypeptide" is used in it broadest sense to refer to a compound of two or more subunit amino acids, amino acid analogs, or other peptidomimetics. The subunits may be linked by peptide bonds or by other bonds, for example ester, ether, etc. As used  
25 herein, the term "amino acid" refers to either natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. A peptide of three or more amino acids is commonly called an oligopeptide if the peptide chain is short. If the peptide chain is long, the peptide is typically called a polypeptide or a protein.

30 "Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter that is operably linked to a coding sequence (e.g., a reporter expression

cassette) is capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter or other control elements need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. For example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting procaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms. An "isolated polynucleotide" molecule is a nucleic acid molecule separate and discrete from the whole organism with which the molecule is found in nature; or a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but having heterologous sequences (as defined below) in association therewith.

Techniques for determining nucleic acid and amino acid "sequence identity" also are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acid-to-amino

acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more sequences (polynucleotide or amino acid) can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences

5 divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-

10 358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, WI) in the "BestFit" utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis

15 Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). A preferred method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages

20 the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the "Match" value reflects "sequence identity."

Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is

25 BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details

30 of these programs can be found at the following internet address:

<http://www.ncbi.nlm.gov/cgi-bin/BLAST>. When claiming sequences relative to

sequences of the present invention, the desired degrees of sequence identity are at least

80%, 85-90%, preferably 92%, more preferably 95%, and even more preferably 98% sequence identity to the reference sequence (i.e., the sequences of the present invention).

Alternatively, the degree of sequence similarity between polynucleotides can be determined by hybridization of polynucleotides under conditions that form stable  
5 duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 80%-85%, preferably at least about 85%-90%, more preferably at least about 90%-95%, and most preferably at least about 95%-98%  
10 sequence identity over a defined length of the molecules, as determined using the methods above. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that  
15 particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; *DNA Cloning, supra*; *Nucleic Acid Hybridization, supra*.

Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the  
20 efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like,  
25 see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity  
30 (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of

5 hybridizing selectively to a target sequence under "moderately stringent" typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-

10 14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985)

15 Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other

20 hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

25

A "vector" is capable of transferring gene sequences to target cells. Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning, and expression

30 vehicles, as well as integrating vectors.

"Nucleic acid expression vector" or "expression cassette" refers to an assembly which is capable of directing the expression of a sequence or gene of interest. The



nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well.

Expression cassettes described herein may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also  
5 include a bacterial origin of replication, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA (*e.g.*, a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (*e.g.*, a SV40 or adenovirus origin of replication).

An "expression cassette" comprises any nucleic acid construct capable of  
10 directing the expression of a gene/coding sequence of interest. Such cassettes can be constructed into a "vector," "vector construct," "expression vector," or "gene transfer vector," in order to transfer the expression cassette into target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

"Luciferase," unless stated otherwise, includes prokaryotic and eukaryotic  
15 luciferases, as well as variants possessing varied or altered optical properties, such as luciferases that produce different colors of light (*e.g.*, Kajiyama, N., and Nakano, E., *Protein Engineering* 4(6):691-693 (1991)).

"Light-generating" is defined as capable of generating light through a chemical reaction or through the absorption of radiation.

20 A "light generating protein" or "light-emitting protein" is a protein capable of generating light in the visible spectrum (between approximately 350 nm and 800 nm). Examples include bioluminescent proteins such as luciferases, *e.g.*, bacterial and firefly luciferases, as well as fluorescent proteins such as green fluorescent protein (GFP).

"Light" is defined herein, unless stated otherwise, as electromagnetic radiation  
25 having a wavelength of between about 300 nm and about 1100 nm.

"Animal" as used herein typically refers to a non-human mammal, including, without limitation, farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens,  
30 turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered.

A "transgenic animal" refers to a genetically engineered animal or offspring of genetically engineered animals. A transgenic animal usually contains material from at least one unrelated organism, such as from a virus, plant, or other animal. The "non-human animals" of the invention include vertebrates such as rodents, non-human  
5 primates, sheep, dogs, cows, amphibians, birds, fish, insects, reptiles, etc. The term "chimeric animal" is used to refer to animals in which the heterologous gene is found, or in which the heterologous gene is expressed in some but not all cells of the animal.

"Analyte" as used herein refers to any compound or substance whose effects (e.g., induction or repression of a specific promoter) can be evaluated using the test  
10 animals and methods of the present invention. Such analytes include, but are not limited to, chemical compounds, pharmaceutical compounds, polypeptides, peptides, polynucleotides, and polynucleotide analogs. Many organizations (e.g., the National Institutes of Health, pharmaceutical and chemical corporations) have large libraries of chemical or biological compounds from natural or synthetic processes, or fermentation  
15 broths or extracts. Such compounds/analytes can be employed in the practice of the present invention.

As used herein, the term "positive selection marker" refers to a gene encoding a product that enables only the cells that carry the gene to survive and/or grow under certain conditions. For example, plant and animal cells that express the introduced  
20 neomycin resistance (Neo<sup>r</sup>) gene are resistant to the compound G418. Cells that do not carry the Neo<sup>r</sup> gene marker are killed by G418. Other positive selection markers will be known to those of skill in the art. Typically, positive selection markers encode products that can be readily assayed. Thus, positive selection markers can be used to determine whether a particular DNA construct has been introduced into a cell, organ or tissue.

25 "Negative selection marker" refers to gene encoding a product which can be used to selectively kill and/or inhibit growth of cells under certain conditions. Non-limiting examples of negative selection inserts include a herpes simplex virus (HSV)-thymidine kinase (TK) gene. Cells containing an active HSV-TK gene are incapable of growing in the presence of gangcylovir or similar agents. Thus, depending on the substrate, some  
30 gene products can act as either positive or negative selection markers.

The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules or chromatids at the site of essentially identical nucleotide

sequences. It is understood that substantially homologous sequences can accommodate insertions, deletions, and substitutions in the nucleotide sequence. Thus, linear sequences of nucleotides can be essentially identical even if some of the nucleotide residues do not precisely correspond or align (see, above).

5       A “knock-out” mutation refers to partial or complete loss of expression of at least a portion the target gene. Examples of knock-out mutations include, but are not limited to, gene-replacement by heterologous sequences, gene disruption by heterologous sequences, and deletion of essential elements of the gene (e.g., promoter region, portions of a coding sequence). A “knock-out” mutation is typically identified by the phenotype  
10 generated by the mutation.

A “single-copy gene” as used herein refers to a gene represented in an organism’s genome only by a single copy at a particular chromosomal locus. Accordingly, a diploid organism has two copies of the gene and both copies occur at the same chromosomal location.

15       A “gene” as used in the context of the present invention is a sequence of nucleotides in a genetic nucleic acid (chromosome, plasmid, etc.) with which a genetic function is associated. A gene is a hereditary unit, for example of an organism, comprising a polynucleotide sequence (e.g., a DNA sequence for mammals) that occupies a specific physical location (a “gene locus” or “genetic locus”) within the  
20 genome of an organism. A gene can encode an expressed product, such as a polypeptide or a polynucleotide (e.g., tRNA). Alternatively, a gene may define a genomic location for a particular event/function, such as the binding of proteins and/or nucleic acids (e.g., phage attachment sites), wherein the gene does not encode an expressed product. Typically, a gene includes coding sequences, such as, polypeptide encoding sequences,  
25 and non-coding sequences, such as, promoter sequences, poly-adenylation sequences, transcriptional regulatory sequences (e.g., enhancer sequences). Many eucaryotic genes have “exons” (coding sequences) interrupted by “introns” (non-coding sequences). In certain cases, a gene may share sequences with another gene(s) (e.g., overlapping genes).

“Isogenic” means two or more organisms or cells that are considered to be  
30 genetically identical. “Substantially isogenic” means two or more organisms or cells wherein, at the majority of genetic loci (e.g., greater than 99.000%, preferably more than 99.900%, more preferably greater than 99.990%, even more preferably greater than

99.999%), there exists genetic identity between the organisms or cells being compared.

In the context of the present invention, two organisms (for example, mice) are considered to be "substantially isogenic" if, for example, inserted transgenes are the primary differences between the genetic make-up of the mice being compared. Further, if, for example, the genetic backgrounds of the mice being compared are the same with the exception that one of the mice has one or several defined mutation(s) (for example, affecting coat color), then these mice are considered to be substantially isogenic. An example of two strains of substantially isogenic mice are C57BL/6 and C57BL/6-Tyr C2j/+.

10 A "pseudogene" as used herein, refers to a type of gene sequence found in the genomes, typically, of eucaryotes, where the sequence closely resembles a known functional gene, but differs in that the pseudogene is non-functional. For example, the pseudogene sequence may contain several stop codons in what would correspond to an open reading frame in the functional gene. Pseudogenes can also have deletions or  
15 insertions relative to their corresponding functional gene. If, for example, in a genome there is a functional gene and a related pseudogene, the functional gene is considered to be a single-copy gene (accordingly, the pseudogene is considered to be single-copy as well).

A "non-essential gene" refers to a gene whose deletion, disruption, elimination, reduction of gene function, or mutation is non-lethal, and does not obviously adversely affect the organisms' ability to mature and reproduce. A "non-essential gene with no phenotype" refers to a non-essential gene whose deletion, disruption, elimination, reduction of gene function or mutation has no deleterious effect on the organism. Typically there are no phenotypically reflected gene dosage effects associated with  
25 modification of a non-essential gene with no phenotype -- for example, deletion, disruption or mutation of both copies of a non-essential gene with no phenotype in a diploid organism has essentially the same effect as deletion, disruption, or mutation of one of the two copies present in the diploid organism. In the context of the present invention, a non-essential gene is typically one whose function has been eliminated (e.g.,  
30 by a deletion mutation) and such elimination of function was non-lethal and the organism developed, matured, and was able to reproduce.

The "native sequence" or "wild-type sequence" of a gene is the polynucleotide sequence that comprises the genetic locus corresponding to the gene, e.g., all regulatory and open-reading frame coding sequences required for expression of a completely functional gene product as they are present in the wild-type genome of an organism. The native sequence of a gene can include, for example, transcriptional promoter sequences, translation enhancing sequences, introns, exons, and poly-A processing signal sites. It is noted that in the general population, wild-type genes may include multiple prevalent versions that contain alterations in sequence relative to each other and yet do not cause a discernible pathological effect. These variations are designated "polymorphisms" or "allelic variations."

By "replacement sequence" is meant a polynucleotide sequence that is substituted for at least a portion of the native or wild-type sequence of a gene.

"Linear vector" or "linearized vector," as used herein, is a vector having two ends. For example, circular vectors, such as plasmids, can be linearized by digestion with a restriction endonuclease that cuts at a single site in the plasmid. Preferably, the targeting vectors described herein are linearized such that the ends are not within the targeting sequences.

### General Overview

The present invention relates to vector constructs and methods of creating transgenic animals to be used, for example, as test systems. Methods of using the animals of the present invention include, but are not limited to, using these animals for studies involving tumor growth and other disease conditions. In the practice of the present invention, transgenic, non-human mammals are constructed where a single-copy, non-essential gene is replaced by a reporter expression cassette, preferably a gene encoding a light-generating protein, such as a luciferase-encoding gene, operably linked to a promoter. A variety of promoters are useful in the practice of the present invention, for example, promoters derived from genes associated with tumorigenesis or angiogenesis. Thus, an exemplary promoter can be one that is associated with proteins induced during tumorigenesis, for instance in the presence of tumor generating compounds or of tumors themselves. In this way, expression of the reporter cassette is induced in the animal when, for example, tumors are present, and progression of the

tumor can be evaluated by non-invasive imaging methods using the whole animal. Another exemplary promoter is one that is derived from a gene associated with angiogenesis. Because the promoter is linked to a reporter such as luciferase, non-invasive monitoring of the progression of angiogenesis is possible.

5        Various forms of the different embodiments of the invention, described herein, may be combined.

Non-invasive imaging and/or detecting of light-emitting conjugates in mammalian subjects was described in U.S. Patent No. 5,650,135, by Contag, et al., issued 22 July 1997. This imaging technology can be used in the practice of the present invention in view of the teachings of the present specification. In the imaging method, the conjugates contain a biocompatible entity and a light-generating moiety. Biocompatible entities include, but are not limited to, small molecules such as cyclic organic molecules; macromolecules such as proteins; microorganisms such as viruses, bacteria, yeast and fungi; eukaryotic cells; all types of pathogens and pathogenic substances; and particles such as beads and liposomes. In another aspect, biocompatible entities may be all or some of the cells that constitute the mammalian subject being imaged, for example, cells carrying the vector constructs of the present invention expressing a reporter expression cassette.

Light-emitting capability is conferred on the biocompatible entities by the conjugation of a light-generating moiety. Such moieties include fluorescent molecules, fluorescent proteins, enzymatic reactions giving off photons and luminescent substances, such as bioluminescent proteins. In the context of the present invention, light emitting capability is typically conferred on target cells by having at least one copy of a light-generating protein, *e.g.*, a luciferase, present. In preferred embodiments, luciferase is operably linked to appropriate control elements which can facilitate expression of a polypeptide having luciferase activity. Substrates of luciferase can be endogenous to the cell or applied to the cell or system (*e.g.*, injection into a transgenic mouse, having cells carrying a luciferase construct, of a suitable substrate for the luciferase, for example, luciferin). The conjugation may involve a chemical coupling step, genetic engineering of a fusion protein, or the transformation of a cell, microorganism or animal to express a light-generating protein.

### Targeting Constructs

The targeting cassettes described herein typically include the following components: (1) a suitable vector backbone; (2) a polynucleotide encoding a light generating protein (3) a promoter operably linked to the light generating protein-  
5 encoding gene, wherein the promoter is heterologous to the light generating protein coding sequences; (4) a sequence encoding a positive selection marker; (5) insertion sites flanking the sequence encoding the positive selection marker and the polynucleotide encoding a light generating protein gene, for insertion of sequences which target a single-copy, non-essential chromosomal gene; and, optionally, (6) a sequence encoding a  
10 negative selection marker. Exemplary targeting constructs are shown in Figures 3B, 5A and 5B and described in Examples 1-3.

Suitable vector backbones generally include an F1 origin of replication; a colE1 plasmid-derived origin of replication; polyadenylation sequence(s); sequences encoding antibiotic resistance (*e.g.*, ampicillin resistance) and other regulatory or control elements.  
15 Non-limiting examples of appropriate backbones include: pBluescriptSK (Stratagene, La Jolla, CA); pBluescriptKS (Stratagene, La Jolla, CA) and other commercially available vectors.

In one aspect of the invention the light generating protein is luciferase. Luciferase coding sequences useful in the practice of the present invention include, but  
20 are not limited to, sequences obtained from *lux* genes (procaryotic genes encoding a luciferase activity) and *luc* genes (eucaryotic genes encoding a luciferase activity). A variety of luciferase encoding genes have been identified including, but not limited to, the following: B.A. Sherf and K.V. Wood, U.S. Patent No. 5,670,356, issued 23 September 1997; Kazami, J., et al., U.S. Patent No. 5,604,123, issued 18 February 1997;  
25 S. Zenno, et al, U.S. Patent No. 5,618,722; K.V. Wood, U.S. Patent No. 5,650,289, issued 22 July 1997; K.V. Wood, U.S. Patent No. 5,641,641, issued 24 June 1997; N. Kajiyama and E. Nakano, U.S. Patent No. 5,229,285, issued 20 July 1993; M.J. Cormier and W.W. Lorenz, U.S. Patent No. 5,292,658, issued 8 March 1994; M.J. Cormier and W.W. Lorenz, U.S. Patent No. 5,418,155, issued 23 May 1995; de Wet, J.R., et al,  
30 *Molec. Cell. Biol.* 7:725-737, 1987; Tatsumi, H.N., et al, *Biochim. Biophys. Acta* 1131:161-165, 1992; and Wood, K.V., et al, *Science* 244:700-702, 1989. Eukaryotic luciferase catalyzes a reaction using luciferin as a luminescent substrate to produce light,

whereas prokaryotic luciferase catalyzes a reaction using an aldehyde as a luminescent substrate to produce light.

Wild-type firefly luciferases typically have an emission maxima at about 550 nm. Numerous variants with differing emission maxima have also been studied. For example, Kajiyama and Nakano (*Protein Eng.* 4(6):691-693, 1991; U.S. Patent No. 5,330,906, issued 19 July 1994) teach five variant firefly luciferases generated by single amino acid changes to the *Luciola cruciata* luciferase coding sequence. The variants have emission peaks of 558 nm, 595 nm, 607 nm, 609 nm and 612 nm. A yellow-green luciferase with an emission peak of about 540 nm is commercially available from Promega, Madison, WI under the name pGL3. A red luciferase with an emission peak of about 610 nm is described, for example, in Contag et al. (1998) *Nat. Med.* 4:245-247 and Kajiyama et al. (1991) *Prot. Eng.* 4:691-693.

Positive selection markers include any gene which a product that can be readily assayed. Examples include, but are not limited to, a hprt gene (Littlefield, J. W., Science 145:709-710 (1964)), a xanthine-guanine phosphoribosyltransferase (gpt) gene, or an adenosine phosphoribosyltransferase (aprt) gene (Sambrook et al., *supra*), a thymidine kinase gene (i.e. "TK") and especially the TK gene of herpes simplex virus (Giphart-Gassler, M. et al., *Mutat. Res.* 214:223-232 (1989)), a nptII gene (Thomas, K. R. et al., *Cell* 51:503-512 (1987); Mansour, S. L. et al., *Nature* 336:348-352 (1988)), or other genes which confer resistance to amino acid or nucleoside analogues, or antibiotics, etc, for example, gene sequences which encode enzymes such as dihydrofolate reductase (DHFR) enzyme, adenosine deaminase (ADA), asparagine synthetase (AS), hygromycin B phosphotransferase, or a CAD enzyme (carbamyl phosphate synthetase, aspartate transcarbamylase, and dihydroorotase). Addition of the appropriate substrate of the positive selection marker can be used to determine if the product of the positive selection marker is expressed, for example cells which do not express the positive selection marker nptII, are killed when exposed to the substrate G418 (Gibco BRL Life Technology, Gaithersburg, MD).

The targeting vector typically contains insertion sites for inserting targeting sequences (*e.g.*, sequences that are substantially homologous to the target sequences in the host genome where integration of the targeting vector/expression cassette is desired). These insertion sites are preferably included such that there are two sites, one site on



either side of the sequences encoding the positive selection marker, light generating protein (e.g., luciferase) and the promoter. Insertion sites are, for example, restriction endonuclease recognition sites, and can, for example, represent unique restriction sites. In this way, the vector can be digested with the appropriate enzymes and the targeting  
5 sequences ligated into the vector.

Optionally, the targeting construct can contain a polynucleotide encoding a negative selection marker. Suitable negative selection markers include, but are not limited to, HSV-tk (see, e.g., Majzoub et al. (1996) *New Engl. J. Med.* 334:904-907 and U.S. Patent No. 5,464,764), as well as genes encoding various toxins including the  
10 diphtheria toxin, the tetanus toxin, the cholera toxin and the pertussis toxin. A further negative selection marker gene is the hypoxanthine-guanine phosphoribosyl transferase (HPRT) gene for negative selection in 6-thioguanine.

Exemplary promoters and single-copy, non-essential genes for use in the vector constructs and methods of the present invention are described below.

15

### Promoters

The targeting constructs and transgenic animals described herein contain a sequence encoding a light generating protein, e.g., luciferase, gene operably linked to a promoter. The promoter may be from the same species as the transgenic animal (e.g.,  
20 mouse promoter used in construct to make transgenic mouse) or from a different species (e.g., human promoter used in construct to make transgenic mouse). The promoter can be derived from any gene of interest. In one embodiment of the present invention, the promoter is derived from a gene whose expression is induced during angiogenesis, for example pathogenic angiogenesis like tumor development. Thus, when a tumor begins  
25 to develop in a transgenic animal carrying a vector construct of the present invention, the promoter is induced and the animal expresses luciferase, which can then be monitored *in vivo*.

Exemplary promoters for use in the present invention are selected such that they are functional in a cell type and/or animal into which they are being introduced.

30 Exemplary promoters include, but are not limited to, promoters obtained from the following mouse genes: vascular endothelial growth factor (VEGF) (VEGF promoter described in U.S. Patent No. 5,916,763; Shima et al. (1996) *J. Bio. Chem.* 271:3877-

3883; sequence available on NCBI under accession number U41383); VEGFR2, also known as Flk-1, (VEGFR-2 promoter described, for example, in Röncke et al. (1996) *Circ. Res.* 79:277-285; Patterson et al. (1995) *J. Bio. Chem.* 270:23111-23118; Kappel et al. (1999) *Blood* 93:4282-4292; sequence available as accession number X89777 of NCBI database); Tie2, also known as Tek (Tie2 promoter described, for example, in Fadel et al. (1998) *Biochem. J.* 338:335-343; Schlaeger et al. (1995) *Develop.* 121:1089-1098; Schlager et al. (1997) *PNAS USA* 94:3058-3063). VEGF is a specific mitogen for EC *in vitro* and a potent angiogenic factor *in vivo*. In a tumorigenesis study, it was shown that VEGF was critical for the initial subcutaneous growth of T-47D breast carcinoma cells transplanted into nude mice, whereas other angiogenic factors, such as, bFGF can compensate for the loss of VEGF after the tumors have reached a certain size (Yoshiji, H., et al., 1997 *Cancer Research* 57: 3924-28). VEGF is a major mediator of aberrant EC proliferation and vascular permeability in a variety of human pathologic situation, such as, tumor angiogenesis, diabetic retinopathy and rheumatoid arthritis (Benjamin LE, et al., 1997 *PNAS* 94: 8761-66; Soker, S., et al., 1998 *Cell* 92: 735-745). VEGF is synthesized by tumor cells *in vivo* and accumulates in nearby blood vessels. Because leaky tumor vessels initiate a cascade of events, which include plasma extravasation and which lead ultimately to angiogenesis and tumor stroma formation, VEGF plays a pivotal role in promoting tumor growth (Dvorak, H.F., et al., 1991 *J Exp Med* 174:1275-8). VEGF expression was upregulated by hypoxia (Shweiki, D., et al., 1992 *Nature* 359: 843-5). VEGF is also upregulated by overexpression of v-Src oncogene (Mukhopadhyay, D., et al., 1995 *Cancer Res.* 15: 6161-5), c-SRC (Mukhopadhyay, D., et al., 1995 *Nature* 375: 577-81), and mutant ras oncogene (Plate, K.H., et al., 1992 *Nature* 359: 845-8). The tumor suppressor p53 downregulates VEGF expression (Mukhopadhyay, D., et al., 1995 *Cancer Res.* 15: 6161-5).

A number of cytokines and growth factors, including PGF and TPA (Grugel, S., et al., 1995 *J. Biological Chem.* 270: 25915-9), EGF, TGF- $\beta$ , IL-1, IL-6 induce VEGF mRNA expression in certain type of cells (Ferrara, N., et al., 1997 *Endocr. Rev.* 18: 4-25). Kaposi's sarcoma-associated herpesvirus (KSHV) encoded a G-protein-coupled receptor, a homolog of IL-8 receptor, can activate JNK/SAPK and p38MAPK and increase VEGF production, thus causing cell transformation and tumorigenicity (Bais, C., et al., *Nature* 1998 391:86-9). VEGF overexpression in skin of transgenic mice

induces angiogenesis, vascular hyperpermeability and accelerated tumor development (Larcher, F., et al., *Oncogene* 1998 17:303-11).

VEGF-B (cDNA sequences available on databases) is a mitogen for EC and may be involved in angiogenesis in muscle and heart (Olofsson, B., et al., 1996 *Proc Natl Acad Sci U S A* 93:2576-81). Shown *in vitro*, binding of VEGF-B to its receptor VEGFR-1 leads to increased expression and activity of urokinase type plasminogen activator and plasminogen activator inhibitor, suggesting a role for VEGF-B in the regulation of extracellular matrix degradation, cell adhesion, and migration (Olofsson, B., et al., 1998 *Proc Natl Acad Sci U S A* 95:11709-14).

10 VEGF-C (see, e.g., U.S. Patent No. 5,916,763 and Shima et al., *supra*) may regulate angiogenesis of lymphatic vasculature, as suggested by the pattern of VEGF-C expression in mouse embryos (Kukk, E., et al., 1996 *Development* 122: 3829-37). Although VEGF-C is also a ligand for VEGFR-2, the functional significance of this potential interaction is unknown. Overexpression of VEGF-C in the skin of transgenic  
15 mice resulted in lymphatic, but not vascular, endothelial proliferation and vessel enlargement, suggesting the major function of VEGF-C is through VEGFR-3 rather than VEGFR-2 (Jeltsch M, et al., 1997 *Science* 276:1423-5). Shown by the CAM assay, VEGF and VEGF-C are specific angiogenic and lymphangiogenic growth factors, respectively (Oh, S.J., et al., (1997) *Devel. Biol.* 188: 96-109). VEGF-C overexpression  
20 in the skin of transgenic mice resulted in lymphatic, but not vascular, endothelial proliferation and vessel enlargement (Jeltsch M, et al., 1997 *Science* 276:1423-5).

VEGF-D (cDNA sequences available on databases) is a mitogen for EC. Given that VEGF-D can also activate VEGFR-3, it is possible that VEGF-D could be involved in the regulation of growth and/or differentiation of lymphatic endothelium (Achen,  
25 M.G., et al., 1998 *Proc Natl Acad Sci U S A* 95: 548-53). VEGF-D is induced by transcription factor c-Fos in mouse (Orlandini, M., 1996 *PNAS* 93: 11675-80).

VEGFR-1 signaling pathway may regulate normal endothelial cell-cell or cell matrix interactions during vascular development, as suggested by the knockout study (Fong, G.H., et al., 1995 *Nature* 376: 65-69). Although VEGFR-1 has a higher affinity  
30 to VEGF than VEGFR-2, it does not transduce the mitogenic signals of VEGF in ECs (Soker, S., et al., 1998 *Cell* 92: 735-745). VEGFR-2 (see, e.g., Rönische et al., Patterson et al., Kappel et al. (1999), *supra*) appears to be the major transducer of VEGF signals in

EC that result in chemotaxis, mitogenicity and gross morphological changes in target cells (Soker, S., et al., 1998 *Cell* 92: 735-745). VEGFR-3 has an essential role in the development of the embryonic cardiovascular system before the emergence of lymphatic vessels, as shown by the knockout study (Dumont, D.J., et al., 1998 *Science* 282: 946-949). Neuropillin-1 (see, e.g., Soker et al. (1998) *Cell* 92:735-745) is a receptor for VEGF165. It can enhance the binding of VEGF165 to VEGFR-2 and VEGF165 mediated chemotaxis (Soker, S., et al., 1998 *Cell* 92: 735-745). Neuropillin1 overexpression in transgenic mice resulted in embryonic lethality. The embryos possessed excess capillaries and blood vessels. Dilated vessels and hemorrhage were also observed (Kitsukawa, T., et al., 1995 *Development* 121: 4309-18).

Further promoters of interest include, but are not limited to, the following. Ang2 is expressed only at predominant vascular remodeling sites, such as ovary, placenta, uterus (Maisonpierre, P.C., et al., 1997 *Science* 277: 55-60). In glioblastoma angiogenesis, Ang2 is found to be expressed in endothelial cells of small blood vessel and capillaries while Ang1 is expressed in glioblastoma tumor cells (Stratmann, A., 1998 *Am J Pathol* 153: 1459-66). Ang2 is up-regulated in bovine microvascular endothelial by VEGF, bFGF, cytokines, hypoxia (Mandriota, S.J., 1998 *Circ Res* 83: 852-9). Ang2 transgenic overexpression disrupts angiogenesis, and is embryonic lethal (Maisonpierre, P.C., et al., 1997 *Science* 277: 55-60). Ang1 is widely expressed, less abundant in heart and liver (Maisonpierre, P.C., et al., 1997 *Science* 277: 55-60). Ang1 is expressed in mesenchymal cells and may up-regulate the expression of Tie2 in the endothelial cells (Suri, C., et al., 1996 *Cell* 87: 1171-1180). Ang1 overexpression in the skin of transgenic mice produces larger, more numerous, and more highly branched vessels (Suri, C., et al., *Science* 1998 282:468-71). Tie2 (see, e.g., Fadel et al.; Schlaeger et al. (1995), and Schlager et al. (1997), *supra*) is endothelial cell specific, up-regulated during wound healing, follicle maturation (Puri, M.C., et al., 1995 *EMBO J* 14: 5884-91) and pathologic angiogenesis (Kaipainen, A., 1994 *Cancer Research* 54: 6571-77), such as, glioblastoma (Stratmann, A., 1998 *Am J Pathol* 153: 1459-66). Tie2 is also expressed in non-proliferating adult endothelium and endothelial cell lines (Dumont, D.J., et al. (1994) *Genes & Develop.* 8:1897-1909). A Tie2 activating mutation causes vascular dysmorphogenesis (Vikkula M, et al., 1996 *Cell* 87: 1181-1190). Tie2 mutant overexpression in transgenic mice is embryonic lethal (Dumont, D.J., et al., *supra*).

Other promoters useful in the practice of the present invention include, by way of example, promoters derived from the sequences encoding the following polypeptide products: PTEN (dual specificity phosphatase); BAI (brain-specific angiogenesis inhibitor); KAI1 (KANGAI 1); catenin beta-1 (cadherin-associated protein, beta); COX2 (PTGS2 cyclooxygenase 2, a.k.a. prostaglandin-endoperoxide synthase 2); MMP2 (72 kDa Type IV-A collagenase); MMP9 (92 kDa type IV-B collagenase); TIMP2 (tissue inhibitor of metalloproteinase 2); and TIMP3 (tissue inhibitor of metalloproteinase 3).

PTEN is a tumor suppressor gene and encodes a protein of 403 amino acids. (Li et al. (1997) *Science* 275:1943-1946; DiCristofano et al. (1998) *Nature Genet.* 19:348-355). Overexpression of PTEN has been shown to inhibit cell migration and it is postulated that this protein may function as a tumor suppressor by negatively regulating cell interactions with the extracellular matrix or by negatively regulating the PI3K/PKB/Akt signaling pathway. (Tamura et al. (1998) *Science* 280:1614-1617; Stambolic et al. (1998) *Cell* 95:29-29). Mutations in PTEN have been detected in cancer cell lines and in the germline of patients having Cowden disease, Lhermitte-Duclos disease and Bannayan-Zonana syndrome (diseases and syndromes which are characterized by hyperplastic/dysplastic changes in the prostate, skin and colon and which are associated with an increased risk of certain cancers, for example, breast cancer, prostate cancer and colon cancer). (Marsh et al. (1998) *Hum. Molec. Genet.* 7:507-515; Marsh et al. (1998) *J. Med. Genet.* 35:881-885; Nelen et al. (1997) *Hum. Molec. Genet.* 6:1383-1387).

BAI1 protein is predicted to be 1,584 amino acids in length and includes an extracellular domain, an intracellular domain and a 7-span transmembrane region similar to that of the secretin receptor. (Nishimori et al. (1997) *Oncogene* 15:245-2150). The extracellular region of BAI1 has a single Arg-Gly-Asp (RGD) motif recognized by integrins and also has five sequences corresponding to the thrombospondin type I (accession number 188060) repeats that can inhibit angiogenesis includes by basic fibroblast growth factor (bFGF, accession number 134920). Shiratsuchi et al. (1997) *Cytogenet. Cell Genet.* 79:103-108, cloned 2 other brain-specific angiogenesis inhibiting genes, designated BAI2 (accession number 602683) and BAI 3 (accession number 602684). Thus, it is postulated that members of this gene family may play a role in suppression of glioblastoma.

KAI1 encodes a 267 amino acid protein which is a member of the leukocyte surface glycoprotein family. The protein has 4 hydrophobic transmembrane domains and 1 large extracellular hydrophilic domain with three potential N-glycosylation sites. (Dong et al. (1995) *Science* 268:884-886). Molecular analysis of KAI1 is described, for example, in Dong et al. (1997) *Genomics* 41:25-32. KAI1 is a tumor metastasis suppressor gene that is capable of inhibiting the metastatic process in experimental animals. Expression of KAI1 is downregulated during tumor progression of prostate, breast, lung, bladder and pancreatic cancers in humans, apparently at the transcriptional or postranscriptional level. Mashimo et al. (1998) *PNAS USA* 95:11307-11311, found that the tumor suppressor gene p53 can directly inactivate the KAI1 gene by interacting with the region 5' to the coding sequence, suggesting a direct relationship between p53 and KAI1.

Catenin beta-1 is an adherens junction (AJ) protein, which are critical for establishing and maintaining epithelial cell layers, for instance during embryogenesis, wound healing and tumor cell metastasis. Molecular analysis, including description of sequence homology to plakoglobin (accession number 173325), homology to the drosophila gene "armadillo" and interactions with Lef1/Tcf DNA binding proteins, is described, for example, in Nollet et al. (1996) *Genomics* 32:413-424; McCrea et al. (1991) *Science* 254:1359-1361 and Korinek et al. (1997) *Science* 275:1784-1787. In addition, studies by Korinek et al., *supra* and Morin et al. (1997) *Science* 275:1787-1790, have indicated that APC (accession number 175100) negatively regulates catenin beta and that regulation of this protein is critical to the tumor suppressive effect of APC. Abnormally high levels of beta-catenin have been detected in certain human melanoma cell lines. (Rubinfeld et al. (1997) *Science* 275:1790-1792. Koch et al. (1999) *Cancer Res.* 59:269-273 report that childhood hepatoblastomas frequently carry a mutated degradation targeting box of the beta-catenin gene. Transgenic mice which express catenin beta under the control of an epidermal promoter undergo de novo hair morphogenesis and eventually these animals develop two types of tumors -- epithelioid cysts and trichofolliculomas. Gat et al. (1998) *Cell* 95:605-614.

COX2 encodes a cyclooxygenase and is a key regulator of prostaglandin synthesis. (Hla et al. (1992) *PNAS USA* 89:7384-7388; Jones et al. (1993) *J. Biol. Chem.* 268:9049-9054). In particular, COX2 is generally considered to be a mediator of

inflammation and overexpression of COX2 in rat epithelial cells results in elevated levels of E-cadherin and Bcl2. (Tsuji & DuBois (1995) *Cell* 83:493-501). In co-cultures of endothelial cells and colon carcinoma cells, cells that overexpress COX2 produce prostaglandins, proangiogenic factors and stimulate both endothelial migration and tube  
5 formation. (Tsuji et al. (1998) *Cell* 93:705-716). Experiments conducted using APC knock-out mice have demonstrated that animals homozygous for a disrupted COX2 locus develop significantly more adenomatous polyps. (Oshima et al. (1996) *Cell* 87:803-809). COX-2 "knock out" mice develop severe nephropathy, are susceptible to peritonitis, exhibit reduced arachidonic acid-induced inflammation and exhibit reduced  
10 indomethacin-induced gastric ulceration. (Morham et al. (1995) *Cell* 83:473-482; Langenbach et al. (1995) *Cell* 83:483-492). Female mice that are deficient in cyclooxygenase 2 exhibit multiple reproductive failures. (Lim et al. (1997) *Cell* 91:197-208).

MMP2 is a metalloproteinase that specifically cleaves type IV collagen. A C-  
15 terminal fragment of MMP2, termed PEX, prevents normal binding to alpha-V/beta-3 and disrupts angiogenesis and tumor growth. (Brooks et al. (1998) *Cell* 92:391-400).

MMP9 is a collagenase secreted from normal skin fibroblasts. MMP9 null mice exhibit an abnormal pattern of skeletal growth plate vascularization and ossification. (Vu et al. (1998) *Cell* 93:411-422).

20 TIMP2 is a collagenase and appears to play a major role in modulating the activity of interstitial collagenase and a number of connective tissue metalloendoproteases. (Stetler-Stevenson et al. (1989) *J. Biol. Chem.* 264:17372-17378). Unlike TIMP1 and TIMP3, TIMP2 is not upregulated by TPA or TGF-beta. (Hammani et al. (1996) *J. Biol. Chem.* 271:25498-25505).

25 TIMP3 (Wilde et al. (1994) *DNA Cell Biol.* 13:711-718) is localized in the extracellular matrix in both its glycosylated and unglycosylated forms. Studies of mutant TIMP3 proteins have demonstrated that C-terminal truncations do not bind to the extracellular matrix. (Langton et al. (1998) *J. Biol. Chem.* 273:16778-16781).

As one of skill in the art will appreciate in view of the teachings of the present  
30 specification, promoter sequences can be easily derived and isolated from known polypeptide sequences or from cDNA or genomic sequences, using method known in the art in view of the teachings herein. An exemplary method of isolating promoter

sequences using cDNA is via a GenomeWalker® kit, commercially available from Clontech (Palo Alto, CA), and described on page 27 of the 1997-1998 Clontech catalog.

### Targeting Sequences: Non-Essential Genes

5 Central to the present invention is the fact that the targeting constructs contain “targeting” sequences (flanking, for example, the light generating protein-encoding sequence and promoter) derived from a single-copy, non-essential gene. These targeting sequences in the construct act via homologous recombination to replace at least a portion of the non-essential gene in the genome with the light-generating protein-encoding (*e.g.*,  
10 luciferase-encoding) sequence operably linked to a promoter.

Non-limiting examples of targeting sequences for use in generating transgenic mice include sequences obtained from or derived from vitronectin, Fos B and galactin 3. A search of Mouse Knockout & Mutation Database (Genome Systems, Inc., St. Louis, MO) can be used to identify genes that have been knocked-out in mice where the  
15 generated knockout mice displayed no obvious defects. The chromosomal locus for all these genes can be used to target promoter-(light generating protein, *e.g.*, luciferase) transgenes similar to what is described in Example 2. Single-copy, non-essential mouse genes identified in this manner include, but are not limited to, the following: Moesin (Msn), Doi Y., et al., J Biol Chem 1999, 274:2315–2321; Plasminogen activator  
20 inhibitor, type II (Planh2) and Planh1, Dougherty K.M., Proc Natl Acad Sci USA 1999, 96:686–691; Protein tyrosine phosphatase, receptor type, B (Ptpb), Elchebly et al. (1999) Science 283:1544–1548; Presenilin 1 (Psen1), Guo Q, et al. (1999) Proc Natl Acad Sci USA, 96:4125–4130; Protein kinase, mitogen-activated 9 (Prkm9) / SAPK/Erk/kinase 2 (Serk2), Kuan CY et al. (1999) Neuron 4:667–676; CD152 antigen  
25 (Cd152) / CD86 antigen (Cd86) / CD80 antigen (Cd80), Mandelbrot DA, et al. (1999) J Exp Med, 189:435–440; Poly (ADP-ribose) polymerase (Adprp), Masutani M, et al. (1999), Proc Natl Acad Sci USA 96:2301–2304; Sodium channel, nonvoltage-gated 1 beta (Scnn1b), Pradervand S, et al. (1999) Proc Natl Acad Sci USA 96:1732–1737; Nuclear receptor coactivator 1 (Ncoa1), Qi C, et al. (1999) Proc Natl Acad Sci USA  
30 96:1585–1590; Decay accelerating factor 1 (Daf1), Sun X, et al. (1999) Proc Natl Acad Sci USA 1999, 96:628–633; Necdin (Ndn), Tsai TF, et al. (1999) Nat Genet 22:15–16; Relaxin (Rln); Zhao L, et al. (1999) Endocrinology 140:445–453; Adenylyl cyclase 8



- (Adcy8), Abdel-Majid RM, et al. (1998) *Nat Genet* 19:289–291; Leukemia inhibitory factor (Lif), Bugga L, et al. (1998) *J Neurobiol* 36:509–524; Lectin, galactose binding, soluble 3 (Lgals3) and Lgals1, Calnot C, et al. (1998) *Dev Dyn* 211:306–313; Urokinase plasminogen activator receptor (Plaur) Carmeliet P, et al. (1998) *J Cell Biol* 140:233–
- 5 245; Nitric oxide synthase 1, neuronal (Nos1), Chao DS, et al. (1998) *J Neurochem* 71:784–789; Homeo box A7 (Hoxa7), Chen F, et al. (1998) *Mech Dev* 77:49–57; Myosin light chain, phosphorylatable, cardiac ventricles (Mylpc) Chen J, et al. (1998) *J Biol Chem* 273:1252–1256; Homeo box B7 (Hoxb7), Chen F, et al. (1998) *Mech Dev* 77:49–57; Nuclear factor of kappa light chain gene enhancer in B-cells inhibitor, alpha and beta (Nfkb1a and Nfkb1b), Cheng JD, et al. (1998) *J Exp Med* 186:1055–1062; Enolase 1, alpha non-neuron (Eno1), Couldrey C, et al. (1998) *Dev Dyn* 212:284–292; Xeroderma pigmentosum, complementation group A (Xpa), De Vries A, et al. (1998) *Exp Eye Res* 1998, 67:53–59; Von Willebrand factor homolog (Vwf), Denis C, et al. (1998) *Proc Natl Acad Sci USA* 95:9524–9529; Lysosomal acid lipase 1 (Lip1), Du H, et al. (1998) *Hum Mol Genet* 7:1347–1354; UNC-5 homolog (*C. elegans*) 3 (Unc5h3), Eisenman LM, et al. (1998) *J Comp Neurol* 394:106–117; Protein phosphatase 1, regulatory (inhibitor) subunit 1B (Ppp1r1b), Fienberg AA, et al. (1998) *Science* 281:838–842; Myelin-associated glycoprotein (Mag) Fujita N, et al. (1998) *J Neurosci* 18:1970–1978; Paraoxonase 1 (Pon1), Furlong CE, et al. (1998) *Neurotoxicology*
- 20 19:645–650; Brain derived neurotrophic factor (Bdnf), Garek RR, et al. (1998) *Laryngoscope* 108:671–678; Neurotrophin 3 (Ntf3), Garek RR, et al. (1998) *Laryngoscope* 108:671–678; Myoglobin (Mb); Garry DJ, et al. (1998) *Nature* 395:905–908; Opioid receptor, mu (Oprm), Gaveriaux-Ruff C, et al. (1998) *Proc Natl Acad Sci USA* 95:6326–6330; Neuropeptide Y (Npy), Hollopeter G, et al. (1998) *Int J Obes Relat*
- 25 *Metab Disord* 22:506–512; Procollagen, type I, alpha 1 (Col1a1) Hormuzdi SG, et al. (1998) *Mol Cell Biol* 18:3368–3375; Centromere autoantigen B (Cenpb), Hudson DF, et al. (1998) *J Cell Biol* 141:309–319; Oculocerebrorenal syndrome of Lowe (ocrl), Janne PA, et al. (1998) *J Clin Invest* 101:2042–2053; arachidonate 12-lipoxygenase (Alox12) Johnson EN, et al. (1998) *Proc Natl Acad Sci USA* 95:3100–3105; H19 fetal liver
- 30 mRNA (H19), Jones BK, et al. (1998) *Genes Dev* 12:2200–2207; Hepatocyte nuclear factor 3 gamma (winged helix transcription factor) (Hnf3g), Kaestner KH, et al. (1998) *Mol Cell Biol* 18:4245–4251; Bone morphogenetic protein 2 (Bmp2) / Bone

- morphogenetic protein 7 (Bmp7) and Bmp5, Katagiri T, et al. (1998) *Dev Genet* 22:340–348; Intercellular adhesion molecule (Icam1), Ley K, et al. (1998) *Circ Res* 83:287–294; Glutamyl aminopeptidase (Enpep), Lin Q, et al. *J Immunol* 1998, 160:4681–4687; Prion protein (Prnp), Lipp HP, et al. *Behav Brain Res* 1998, 95:47–54; RAB3A, member RAS oncogene family (Rab3a), Lonart G, et al. *Neuron* 1998, 21:1141–1150; Potassium voltage gated channel, shaker related subfamily, member 4 (Kcna4), London B, et al., *J Physiol (Lond)* 1998, 509:171–182; Apurinic/aprimidinic endonuclease (Apex), Ludwig DL, et al. *Mutat Res* 1998, 409:17–29; T-cell receptor gamma, variable 5 (Tcrg-V5), Mallick-Wood CA, et al. *Science* 1998, 279:1729–1733; Nuclear, factor, erythroid derived 2, like 2 (Nfe2l2), Martin F, et al. *Blood* 1998, 91:3459–3466; Interleukin 13 (Il13), McKenzie GJ, et al. *Curr Biol* 1998, 8:339–340; Sorbitol dehydrogenase 1 (Sdh1), Ng T, et al. *Diabetes* 1998, 47:961–966; Guanine nucleotide binding protein, alpha 11 (Gna11), Offermanns S, et al. *EMBO J* 1998, 17:4304–4312; Estrogen receptor alpha (Estra), Ogawa S, et al. *Endocrinology* 1998, 139:5070–5081; Integrin beta 2 (Itgb2), Intracellular adhesion molecule (Icam1) and CD34 antigen, Oliveira-dos-Santos AJ, et al. *Eur J Immunol* 1998, 28:2882–2892; Angiotensin receptor 1b (Agtr1b), Oliverio MI, et al. *Proc Natl Acad Sci USA* 1998, 95:15496–15501; Complement factor B (factor B), Pekna M, et al. *Scand J Immunol* 1998, 47:375–380; Centromere autoantigen B (Cenpb), Perez-Castro AV, et al. *Dev Biol* 1998, 201:135–143; Procollagen, type V, alpha 2 (Col5a2) / Fibrillin 1 (Fbn1), Phelps RG, et al. *Mol Med* 1998, 4:356–360; Plasminogen activator inhibitor, type 1 (Planh1), Pinsky DJ, et al. *J Clin Invest* 1998, 102:919–928, Carmeliet P, et al. *J Clin Invest* 1993, 92:2756–2760; Placentae and embryos oncofetal gene (Pem), Pitman JL, et al. *Dev Biol* 1998, 202:196–214; Postmeiotic segregation increased 1 (*S. cerevisiae*) (Pms1), Prolla TA, et al. *Nat Genet* 1998, 18:276–278; Prion protein, structural locus (Prn-p), Prusiner SB, et al. *Proc Natl Acad Sci USA* 1998, 95:10608–10612, Lledo P-M, et al. *Proc Natl Acad Sci U S A* 1996, 93:2403–2407, Sailer A, et al. *Cell* 1994, 77:967–968, Weissmann C, et al. *Philos Trans R Soc Lond [Biol]* 1994, 343:431–433, Bueler H, et al. *Cell* 1993, 73:1337–1347, Weissmann C, et al. *Intervirology* 1993, 35:164–175; NAD (P)H:quinone oxidoreductase, Radjendirane V, et al. *J Biol Chem* 1998, 273:7382–7389; Alpha tropomyosin (Tpm1), Rethinasamy P, et al. *Circ Res* 1998, 82:116–123; Goosecoid and Goosecoid-like (Gsc1), Wakamiya M, et al. *Hum Mol Genet* 1998, 7:1835–1840, Saint-Pore B, et al. *Hum Mol Genet* 1998,

- 7:1841–1849; Schlafen 1 (Slfn1); Schwarz DA, et al. *Immunity* 1998, 9:657–668; Nuclear factor, erythroid derived 2, ubiquitous (Mafk), Shavit JA, et al. *Genes Dev* 1998, 12:2164–2174; Microphthalmia-associated transcription factor (Mitf), Smith SB, et al. *Exp Eye Res* 1998, 66:403–410; Bone morphogenetic protein 6 (Bmp6), Solloway MJ, et al. *Dev Genet* 1998, 22:321–339; Phosphatidylinositol glycan, class A (Piga), Takahama Y, et al. *Eur J Immunol* 1998, 28:2159–2166; Paired-related homeobox 2 (Prx2), Berge D, et al. *Development* 1998, 125:3831–3842; Prostaglandin E receptor EP1 subtype (Ptgerp1), Ushikubi F, et al. *Nature* 1998, 395:281–284; Immunoglobulin kappa chain complex (Igk), van der Stoep N, et al. *Immunity* 1998, 8:743–750; Adenine  
 10 phosphoribosyl transferase (Aprt), Van Sloun PP, et al. *Nucleic Acids Res* 1998, 26:4888–4894; Microtubule associated protein 4 (Mtap4), Voss AK, et al. *Dev Dyn* 1998, 212:258–266; ; 3-hydroxy-3-methylglutaryl-coenzyme A lyase (Hmgcl), Wang SP, et al. *Hum Mol Genet* 1998, 7:2057–2062; Fibroblast growth factor receptor 4 (Fgfr4), Weinstein M, et al. *Development* 1998, 125:3615–3623; Hepsin (Hpn), Wu Q,  
 15 et al. *J Clin Invest* 1998, 101:321–326; Small inducible cytokine A11 (Scya11), Yang T, et al. *Blood* 1998, 92:3912–3923; Small nuclear ribonucleoprotein N (Snrpn), Yang T, et al. *Nat Genet* 1998, 19:25–31; DNA fragmentation factor, alpha subunit (Dffa), Zhang J, et al. *Proc Natl Acad Sci USA* 1998, 95:12480–12485; Early growth response 1 (Egr1), Zheng D, et al. *Neuroscience* 1998, 83:251–258; Early growth response 1 (Egr1) /  
 20 Hormone receptor (Hmr), Zheng D, et al. *Neuroscience* 1998, 83:251–258; Hemochromatosis (Hfe), Zhou XY, et al. *Proc Natl Acad Sci USA* 1998, 95:2492–2497; Alpha tropomyosin (Tpm1), Blanchard EM, et al. *Circ Res* 1997, 81:1005–1010; tRNA phosphoserine (Trsp), Bosl MR, et al. *Proc Natl Acad Sci U S A* 1997, 94:5531–5534; Angiotensin receptor 1b (Agtr1b), Chen X, et al. *Am J Physiol* 1997, 272:F299–F304;  
 25 Xeroderma pigmentosum, complementation group C (Xpc), Cheo DL, et al. *Mut Res* 1997, 374:1–9; B cell leukemia/lymphoma 6 (Bcl6), Dent AL, et al. *Science* 1997, 276:589–592; Fumarylacetoacetate hydrolase (Fah) / 4-hydroxyphenylpyruvic acid dioxygenase (Hpd), Endo F, et al. *J Biol Chem* 1997, 272:24426–24432; N-methylpurine-DNA glycosylase (Mpg), Engelward BP, et al. *Proc Natl Acad Sci USA*  
 30 1997, 94:13087–13092; Interleukin 1 receptor, type 1 (Il1r1), Glaccum MB, et al. *J Immunol* 1997, 159:3364–3371; N-methylpurine-DNA glycosylase (Mpg), Hang B, et al. *Proc Natl Acad Sci USA* 1997, 94:12869–12874; Gamma-aminobutyric acid (GABA-

- A) receptor, subunit alpha 6 (Gabra6), Homanics GE, et al. *Mol Pharmacol* 1997, 51:588–596; Superoxide dismutase 2, mitochondrial (Sod2), Huang TT, et al. *Arch Biochem Biophys* 1997, 344:424–432; Interleukin 11 receptor, alpha chain 1 (Il11ra1), Nandurjar HH, et al. *Blood* 1997, 90:2148–2159; Alkaline phosphatase 5 (Akp5),
- 5 Narisawa S, et al. *Dev Dyn* 1997, 208:432–446; GATA-binding protein 4 (Gata4), Narita N, et al. *Development* 1997, 124:3755–3764; Lymphocyte protein tyrosine kinase (Lck) / Fyn protooncogene (Fyn); Page ST, et al. *Eur J Immunol* 1997, 27:554–562; P glycoprotein 3 (Pgy3), Schinkel AH, et al. *Proc Natl Acad Sci U S A* 1997, 94:4028–4033; P glycoprotein 1 (Pgy1) / P glycoprotein 3 (Pgy3), Schinkel AH, et al., *Proc Natl*
- 10 *Acad Sci U S A* 1997, 94:4028–4033; Creatine kinase, mitochondrial 1, ubiquitous (Ckmt1), Steeghs K, et al. *J Neurosci Methods* 1997, 71:29–41; T cell receptor gamma, variable 4 (Tcrg-V4), Sunaga S, et al. *J Immunol* 1997, 158:4223–4228; Ia-associated invariant chain (p31 form) (Ii), Takaesu NT, et al. *J Immunol* 1997, 158:187–199; Solute carrier family 18 (vesicular monoamine), member 2 (Slc18a2), Takahashi N, et al. *Proc*
- 15 *Natl Acad Sci USA* 1997, 94:9938–9943; Matrix metalloproteinase 7 (Mmp7), Wilson CL, et al. *Proc Natl Acad Sci U S A* 1997, 94:1402–1407; Formin (Fmn), Wynshaw-Boris A, et al. *Mol Med* 1997, 3:372–384; Synaptophysin (Syp), Arrandale JM, et al. *J Biol Chem* 1996, 271:21353–21358; Transformation related protein 53 (Trp53), Boehme SA, et al. *J Immunol* 1996, 156:4075–4078; Neuronal nitric oxide synthase (Nos1),
- 20 Burnett AL, et al. *Mol Medicine* 1996, 2:288–296; Eph receptor A2 (Epha2), Chen J, et al. *Oncogene* 1996, 12:979–988; Urokinase plasminogen activator receptor (Plaur); Dewerchin M, et al. *J Clin Invest* 1996, 97:870–878; Growth differentiation factor 9 (Gdf9), Dong J, et al. *Nature* 1996, 383:531–535; Externally regulated phosphatase (Ptpn16), Dorfman K, et al. *Oncogene* 1996, 13:925–931; Tenascin C (Tnc), Forsberg E,
- 25 et al. *Proc Natl Acad Sci U S A* 1996, 93:6594–6599; Integrin alpha 1 (Itga1), Gardner H, et al. *Dev Biol* 1996, 175:301–313; FBJ osteosarcoma oncogene B (Fosb), Gruda MC, et al. *Oncogene* 1996, 12:2177–2185; Breast cancer 1 (Brca1), Hakem R, et al. *Cell* 1996, 85:1009–1023; Megakaryocyte-associated tyrosine kinase (Matk), Hamaguchi I, et al. *Biochem Biophys Res Commun* 1996, 224:172–179; Apolipoprotein B editing
- 30 complex 1 (Apobec1), Hirano K-I, et al. *J Biol Chem* 1996, 271:9887–9890; Carboxyl ester lipase, Howles PN, et al. *J. Biol. Chem.* 1996, 271:7196–7202; Nuclear factor, erythroid derived 2, ubiquitous (Nfe2u), Kotkow KJ, Orkin SH. *Proc Natl Acad Sci U S*

- A 1996, 93:3514–3518; Retinoid X receptor gamma (Rxrg), Krezel W, et al. *Proc Natl Acad Sci U S A* 1996, 93:9010–9014; Early growth response 1 (Egr1), Lee SL, et al. *Mol Cell Biol* 1996, 16:4566–4572; Adrenergic receptor, alpha 2b (Adra2b), Link RE, et al. *Science* 1996, 273:803–805; Angiotensin receptor 1a (Agtr1a), Matsusaka T, et al. *J Clin Invest* 1996, 98:1867–1877; Interleukin 3 receptor, beta chain 1 (Il3rb2), Nicola NA, et al. *Blood* 1996, 87:2665–2674; Transformation related protein 53 (Trp53), Ohashi M, et al. *Jpn J Cancer Res* 1996, 87:696–701; Leukemia-associated gene (Lag), Schubart U K, et al. *J Biol Chem* 1996, 271:14062–14066; Glutathione peroxidase 1 (Gpx1), Spector A, et al. *Exp Eye Res* 1996, 62:521–540; Arachidonate 12-lipoxygenase, leukocyte (Alox12l), Sun D, Funk CD. *J Biol Chem* 1996, 271:24055–24062; Complement component 3 (C3), Sylvestre D, et al. *J Exp Med* 1996, 184:2385–2392; CD30 antigen (Cd30), Texido G, et al. *Eur J Immunol* 1996, 26:1966–1969; Fc receptor, IgE, low affinity II, alpha polypeptide (Fcer2a), Immunoglobulin heavy chain 5 (delta-like heavy chain) (Igh-5), Interleukin 4 (IL4) and Terminal-deoxynucleotidyl transferase, Texido G, et al. *Eur J Immunol* 1996, 26:1966–1969; Apolipoprotein A-II (Apoa2), Weng W, Breslow JL. *Proc Natl Acad Sci U S A* 1996, 93:14788–14794; Amyloid beta (A4) precursor protein (App), Zheng H, et al. *Ann N Y Acad Sci* 1996, 777:421–426; cAMP responsive element binding protein 1 (Creb1), Blendy JA, et al. *Brain Res* 1995, 681:8–14; Bradykinin receptor, beta 2 (Bdkrb2), Borkowski JA, et al. *J Biol Chem* 1995, 270:13706–13710; Growth factor response protein (Gfrp), Crawford PA, et al. *Mol Cell Biol* 1995, 15:4331–4336; Ciliary neurotrophic factor (Cntf), de Chiara TM, et al. *Cell* 1995, 83:313–322; Cyclin dependent kinase inhibitor 1A (P21) (Cdkn1a); Deng C, et al. *Cell* 1995, 82:675–684; Granzyme A (Gzma), Ebnet K, et al. *EMBO J* 1995, 14:4230–4239; Very low density lipoprotein receptor (Vldlr), Frykman PK, et al. *Proc Natl Acad Sci U S A* 1995, 92:8453–8457; Apolipoprotein E (ApoE) / Apolipoprotein A-I (Apoa1), Goodrum JF, et al. *J Neurobiol* 1995, 64:408–416; Nitric oxide synthase 1, neuronal (Nos1), Ichinose F, et al. *Anesthesiology* 1995, 83:101–108; Nitric oxide synthase 2, inducible, macrophage (Nos2), Laubach VE, et al. *Proc Natl Acad Sci U S A* 1995, 92:10688–10692; Peroxisome proliferator activated receptor alpha (Ppara), Lee SS-T, et al. *Mol Cell Biol* 1995, 15:3012–3022; Growth factor response protein (Gfrp), Lee SL, et al. *Science* 1995, 269:532–535; H19 fetal liver mRNA (H19) / Insulin-like growth factor 2 (Igf2), Leighton PA, et al. *Nature* 1995, 375:34–39; Retinoic

- acid receptor beta (RAR-beta), Luo J, et al. *Mech Dev* 1995, 53:61-71; Metallothionein 1 (Mt1) / Metallothionein 2 (Mt2), Philcox JC, et al. *Biochem J* 1995, 308:543-546; Heme oxygenase (decycling) 2 (Hmox2), Poss KD, et al. *Neuron* 1995, 15:867-873; H1-0 histone (H1fv), Sirotkin AM, et al. *Proc Natl Acad Sci U S A* 1995, 92:6434-6438;
- 5 Creatine kinase, mitochondrial 1, ubiquitous (Ckmt1), Steeghs K, et al. *Biochim Biophys Acta* 1995, 1230:130-138; Tenascin C (Tnc), Steindler DA, et al. *J Neurosci* 1995, 15:1971-1983; Ia-associated invariant chain (Ii), Takaesu NT, et al. *Immunity* 1995, 3:385-396; Neuroblastoma ras oncogene (Nras), Umanoff H, et al. *Proc Natl Acad Sci USA* 1995, 92:1709-1713; Receptor-associated protein of the synapse, 43 kDa (Rapsn),
- 10 Willnow TE, et al. *Proc Natl Acad Sci USA* 1995, 92:4537-4541; Vitronectin (Vtn), Zheng X, et al. *Proc Natl Acad Sci U S A* 1995, 92:12426-12430; Preproacrosin (Acr), Baba T, et al. *J Biol Chem* 1994, 269:31845-31849; Vimentin (Vim), Colucci GE, et al. *Cell* 1994, 79:679-694; Tumor necrosis factor receptor superfamily, member 1b (Tnfrsf1b), Erickson SL, et al. *Nature* 1994, 372:560-563; Cellular retinoic acid binding
- 15 protein I (Crabp1), Gorry P, et al. *Proc Natl Acad Sci USA* 1994, 91:9032-9036; cAMP responsive element binding protein 1 (Creb1), Hummler F, et al. *Proc Natl Acad Sci USA* 1994, 91:5647-5651; Pore forming protein (Pfp), Kagi D, et al. *Nature* 1994, 369:31-37; Src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sit (Srms), Kohmura N, et al. *Mol Cell Biol* 1994, 14:6915-6925; CD3
- 20 polypeptide zeta (Cd3z) / Solute carrier family 22, member 1 (Slc22a1), Koyasu S, et al. *EMBO J* 1994, 13:784-797; Pore forming protein (Pfp), Lowin B, et al. *Proc Natl Acad Sci USA* 1994, 91:11571-11575; Retinoic acid receptor, beta (Rarb), Retinoic acid receptor beta2 (RARbeta2), Mendelsohn C, et al. *Dev Biol* 1994, 166:246-258; Transthyretin (Ttr), Palha JA, et al. *J Biol Chem* 1994, 269:33135-33139; Procollagen,
- 25 type X, alpha 1 (Col10a1), Rosati R, et al. *Nature Genet* 1994, 8:129-135; P glycoprotein 3 (Pgy3), Schinkel AH, et al. *Cell* 1994, 77:491-502; Yamaguchi sarcoma viral (v-yes) oncogene homolog (Yes), Stein PL, et al. *Genes Dev* 1994, 8:1999-2007; Fc receptor, IgE, high affinity II, alpha polypeptide (Fcer2a), Stief A, et al. *J Immunol* 1994, 152:3378-3390; Pore forming protein (Pfp), Walsh CM, et al. *Proc Natl Acad Sci*
- 30 *USA* 1994, 91:10854-10858; CD2 antigen (Cd2), Evans CF, et al. *J Immunol* 1993, 151:6259-6264; Mannose-6-phosphate receptor, cation dependent (M6pr), Koster A, et al. *EMBO J* 1993, 12:5219-5223; Retinoic acid receptor, alpha (Rara), Li E, et al. *Proc*

- Natl Acad Sci USA 1993, 90:1590–1594; Lufkin T, et al. Proc Natl Acad Sci USA 1993, 90:7225–7229; Retinoic acid receptor, gamma (Rarg), Lohnes D, et al. Cell 1993, 73:643–658; Tumor necrosis factor receptor 1 (TNF-R-1) (Tnfr1), Pfeffer K, et al. Cell 1993, 73:457–467; Lectin, galactose binding, soluble I (Lgals1), Poirier F, Robertson EJ.
- 5 Development 1993, 119:1229–1236; Synapsin I (Syn1), Rosahl TW, et al. Cell 1993, 75:661–670; Tumor necrosis factor receptor 1 (Tnfr1), Rothe J, et al. Nature 1993, 364:798–802; Beta-2 microglobulin (B2m), Correa I, et al. Proc Natl Acad Sci USA 1992, 89:653–657; CD2 antigen (Cd2), Killeen N, et al. EMBO J 1992, 11:4329–4336; Apolipoprotein E (Apoe), Piedrahita JA, et al. Proc Natl Acad Sci USA 1992, 89:4471–
- 10 4475; Myogenic differentiation 1 (Myod1), Rudnicki MA, et al. Cell 1992, 71:383–390; Tenascin C (Tnc), Saga Y, et al. Genes Dev 1992, 6:1821–1831; Beta-2 microglobulin (B2m), Sanjuan N, et al. J Virol 1992, 66:4587–4590; Neuroblastoma myc-related oncogene 1 (Nmyc1), Stanton BR, et al. Genes Dev 1992, 6:2235–2247; and Hemoglobin alpha chain complex (Hba), Popp RA, et al. Genetics 1983, 105:157–167.
- 15       Some preferred single-copy, non-essential genes with no phenotypes of the present invention include, but are not limited to, the following: Moesin (Msn), Doi Y., et al., J Biol Chem 1999, 274:2315–2321; Plasminogen activator inhibitor, type II (Planh2) and Planh1, Dougherty K.M., Proc Natl Acad Sci USA 1999, 96:686–691; Nuclear receptor coactivator 1 (Ncoa1), Qi C, et al. (1999) Proc Natl Acad Sci USA 96:1585–
- 20 1590; Nuclear factor of kappa light chain gene enhancer in B-cells inhibitor, alpha and beta (Nfkb1a and Nfkb1b), Cheng JD, et al. (1998) J Exp Med 6:1055–1062; H19 fetal liver mRNA (H19), Jones BK, et al. (1998) Genes Dev 12:2200–2207; Prion protein (Prnp), Lipp HP, et al. Behav Brain Res 1998, 95:47–54; Centromere autoantigen B (Cenpb), Perez-Castro AV, et al. Dev Biol 1998, 201:135–143; Placentae and embryos
- 25 oncofetal gene (Pem), Pitman JL, et al. Dev Biol 1998, 202:196–214; Externally regulated phosphatase (Ptpn16), Dorfman K, et al. Oncogene 1996, 13:925–931; Transformation related protein 53 (Trp53), Ohashi M, et al. Jpn J Cancer Res 1996, 87:696–701; H1-0 histone (H1fv), Sirotkin AM, et al. Proc Natl Acad Sci U S A 1995, 92:6434–6438; Creatine kinase, mitochondrial 1, ubiquitous (Ckmt1), Steeghs K, et al.
- 30 Biochim Biophys Acta 1995, 1230:130–138; Neuroblastoma ras oncogene (Nras), Umanoff H, et al. Proc Natl Acad Sci USA 1995, 92:1709–1713; Vitronectin (Vtn), Zheng X, et al. Proc Natl Acad Sci U S A 1995, 92:12426–12430; Vimentin (Vim),

Colucci GE, et al. Cell 1994, 79:679–694; Cellular retinoic acid binding protein I (Crabp1), Gorry P, et al. Proc Natl Acad Sci USA 1994, 91:9032–9036; Retinoic acid receptor beta2 (RARbeta2), Mendelsohn C, et al. Dev Biol 1994, 166:246–258; Retinoic acid receptor, alpha (Rara), Li E, et al. Proc Natl Acad Sci USA 1993, 90:1590–1594,  
 5 Lufkin T, et al. Proc Natl Acad Sci USA 1993, 90:7225–7229; Lectin, galactose binding, soluble 1 (Lgals1), Poirier F, Robertson EJ. Development 1993, 119:1229–1236; Myogenic differentiation 1 (Myod1), Rudnicki MA, et al. Cell 1992, 71:383–390; and Tenascin C (Tnc), Saga Y, et al. Genes Dev 1992, 6:1821–1831.

In view of the guidance of the present specification, one of ordinary skill in the  
 10 art can select similar, suitable, single-copy, non-essential genes in mice and other cell types/organisms.

#### Assembly of Targeting Cassettes

The targeting cassettes described herein can be constructed utilizing  
 15 methodologies known in the art of molecular biology (see, for example, Ausubel or Maniatis) in view of the teachings of the specification. As described above, the targeting constructs are assembled by inserting, into a suitable vector backbone, polynucleotides encoding a reporter, such as a light-generating protein, *e.g.*, a luciferase gene, operably linked to a promoter of interest; a sequence encoding a positive selection marker; and,  
 20 optionally a sequence encoding a negative selection marker. In addition, the targeting cassette contains insertion sites such that sequences targeting a single-copy, non-essential gene can be readily inserted to flank the sequence encoding positive selection marker and luciferase-encoding sequence.

A preferred method of obtaining polynucleotides, suitable regulatory sequences  
 25 (*e.g.*, promoters) is PCR. General procedures for PCR as taught in MacPherson et al., PCR: A PRACTICAL APPROACH, (IRL Press at Oxford University Press, (1991)). PCR conditions for each application reaction may be empirically determined. A number of parameters influence the success of a reaction. Among these parameters are annealing temperature and time, extension time, Mg<sup>2+</sup> and ATP concentration, pH, and the relative  
 30 concentration of primers, templates and deoxyribonucleotides. Exemplary primers are described below in the Examples. After amplification, the resulting fragments can be



detected by agarose gel electrophoresis followed by visualization with ethidium bromide staining and ultraviolet illumination.

In one embodiment, PCR can be used to amplify fragments from genomic libraries. Many genomic libraries are commercially available. Alternatively, libraries  
5 can be produced by any method known in the art. Preferably, the organism(s) from which the DNA is has no discernible disease or phenotypic effects. This isolated DNA may be obtained from any cell source or body fluid (*e.g.*, ES cells, liver, kidney, blood cells, buccal cells, cerviovaginal cells, epithelial cells from urine, fetal cells, or any cells present in tissue obtained by biopsy, urine, blood, cerebrospinal fluid (CSF), and tissue  
10 exudates at the site of infection or inflammation). DNA is extracted from the cells or body fluid using known methods of cell lysis and DNA purification. The purified DNA is then introduced into a suitable expression system, for example a lambda phage.

Another method for obtaining polynucleotides, for example, short, random nucleotide sequences, is by enzymatic digestion. As described below in the Examples,  
15 short DNA sequences generated by digestion of DNA from vectors carrying genes encoding luciferase (yellow green or red).

Polynucleotides are inserted into vector genomes using methods known in the art. For example, insert and vector DNA can be contacted, under suitable conditions, with a restriction enzyme to create complementary or blunt ends on each molecule that can pair  
20 with each other and be joined with a ligase. Alternatively, synthetic nucleic acid linkers can be ligated to the termini of a polynucleotide. These synthetic linkers can contain nucleic acid sequences that correspond to a particular restriction site in the vector DNA. Other means are known and, in view of the teachings herein, can be used.

25 The final constructs can be used immediately (*e.g.*, for introduction into ES cells), or stored frozen (*e.g.*, at -20°C) until use. Preferably, the constructs are linearized prior to use, for example by digestion with suitable restriction endonucleases.

### Transgenic Animals

The targeting constructs containing the light generating protein coding sequences (e.g., luciferase genes) are introduced into a pluripotent cell (e.g., ES cell, Robertson, E. J., In: Current Communications in Molecular Biology, Capecchi, M. R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989), pp. 39-44). Suitable ES cells may be derived or isolated from any species or from any strain of a particular species. Although not required, the pluripotent cells are typically derived from the same species as the intended recipient. ES cells may be obtained from commercial sources, from International Depositories (e.g., the ATCC) or, alternatively, may be obtained as described in Robertson, E. J., *supra*. Examples of clonally-derived ES cells lines include 129/SVJ ES cells, RW-4 and C57BL/6 ES cells (Genome Systems, Inc.).

ES cells are cultured under suitable conditions, for example, as described in Ausubel et al., section 9.16, *supra*. Preferably, ES cells are cultured on stromal cells (such as STO cells (especially SNC4 STO cells) and/or primary embryonic fibroblast cells) as described by E. J. Robertson, *supra*, pp 71-112. Culture media preferably includes leukocyte inhibitory factor ("lif") (Gough, N. M. et al., Reprod. Fertil. Dev. 1:281-288 (1989); Yamamori, Y. et al., Science 246:1412-1416 (1989), which appears to help keep the ES cells from differentiating in culture. Stromal cells transformed with the gene encoding lif can also be used.

The targeting constructs are introduced into the ES cells by any method which will permit the introduced molecule to undergo recombination at its regions of homology, for example, micro-injection, calcium phosphate transformation, or electroporation (Toneguzzo, F. et al., Nucleic Acids Res. 16:5515-5532 (1988); Quillet, A. et al., J. Immunol. 141:17-20 (1988); Machy, P. et al., Proc. Natl. Acad. Sci. (U.S.A.) 85:8027-8031 (1988)). The construct to be inserted into the ES cell must first be in the linear form. Thus, if the knockout construct has been inserted into a vector as described above, linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence. If the ES cells are to be electroporated to insert the construct, the ES cells and construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer's guidelines for use. After electroporation, the ES cells are typically allowed to recover under suitable incubation conditions. The cells

are then cultured under conventional conditions, as are known in the art, and screened for the presence of the construct.

Screening and selection of those cells into which the targeting construct has been integrated can be achieved using the positive selection marker and/or the negative  
5 selection marker in the construct. In preferred embodiments, the construct contains both positive and negative selection markers. In one aspect, methods which rely on expression of the selection marker are used, for example, by adding the appropriate substrate to select only those cells which express the product of the positive selection marker or to eliminate those cells expressing the negative selection marker. For  
10 example, where the positive selection marker encodes neomycin resistance, G418 is added to the transformed ES cell culture media at increasing dosages. Similarly, where the negative selection marker is used, a suitable substrate (*e.g.*, gancyclovir if the negative selection marker encodes HSV-TK) is added to the cell culture. Either before or after selection using the appropriate substrate, the presence of the positive and/or  
15 negative selection markers in a recipient cell can also be determined by others methods, for example, hybridization, detection of radiolabelled nucleotides, PCR and the like. In preferred embodiments, cells having integrated targeting constructs are first selected by adding the appropriate substrate for the positive and/or negative selection markers. Cells that survive the selection process are then screened by other methods, such as PCR or  
20 Southern blotting, for the presence of integrated sequences.

After suitable ES cells containing the construct in the proper location have been identified, the cells can be inserted into an embryo, preferably a blastocyst. The blastocysts are obtained by perfusing the uterus of pregnant females. In one embodiment, the blastocysts are obtained from, for example, the FVB/N strain of mice and the ES cells  
25 are obtained from, for example, the C57BL/6 strain of mice. Suitable methods for accomplishing this are known to the skilled artisan, and are set forth by, *e.g.*, Bradley et al., (1992) *Biotechnology*, 10:534-539. Insertion into the embryo may be accomplished in a variety of ways known to the skilled artisan, however a preferred method is by microinjection. For microinjection, about 10-30 ES cells are collected into a micropipet  
30 and injected into embryos that are at the proper stage of development to permit integration of the foreign ES cell containing the construct into the developing embryo.

The suitable stage of development for the embryo used for insertion of ES cells is species dependent, in mice it is about 3.5 days.

While any embryo of the right stage of development is suitable for use, it is preferred that blastocysts are used. In addition, preferred blastocysts are male and, 5 furthermore, preferably have genes encoding a coat color that is different from that encoded by the genes ES cells. In this way, the offspring can be screened easily for the presence of the knockout construct by looking for mosaic coat color (indicating that the ES cell was incorporated into the developing embryo). Thus, for example, if the ES cell line carries the genes for black fur, the blastocyst selected will carry genes for white or 10 brown fur.

After the ES cell has been introduced into the blastocyst, the blastocyst is typically implanted into the uterus of a pseudopregnant foster mother for gestation. Pseudopregnant females are prepared by mating with vasectomized males of the same species and successful implantation usually must occur within about 2-3 days of mating.

15 Offspring are screened initially for mosaic coat color where the coat color selection strategy has been employed. Southern blots and/or PCR may also be used to determine the presence of the sequences of interest. Mosaic (chimeric) offspring are then bred to each other to generate homozygous animals. Homozygotes and heterozygotes may be identified by Southern blotting of equivalent amounts of genomic DNA from 20 mice that are the product of this cross, as well as mice that are known heterozygotes and wild type mice. Alternatively, Northern blots can be used to probe the mRNA to identify the presence or absence of transcripts encoding either the replaced gene, the light generating protein coding sequence (e.g., luciferase gene), or both. In addition, Western blots can be used to assess the level of expression of the luciferase protein with an 25 antibody against the luciferase gene product. Finally, in situ analysis (such as fixing the cells and labeling with antibody) and/or FACS (fluorescence activated cell sorting) analysis of various cells from the offspring can be conducted using suitable (e.g., anti-luciferase) antibodies to look for the presence or absence of the targeting construct.

In one embodiment of the present invention, the animals are from the C57BL/6 30 mouse strain. This strain develops a variety of tumors and has been used to develop a number of tumor cells lines, for example, B16 melanoma cells (including, B16F10, B16D5, and B16F1), Lewis lung carcinoma cells (including, LLC, LLC-h59), T241

mouse fibrosarcoma cells, RM-1 and pTC2 mouse prostate cancer cells, and MCA207 mouse sarcoma cells. These cell lines have been extensively used for *in vivo* tumor biology studies after injection into C57BL/6 mice. The generated targeted transgenic mice in the Examples are in C57BL/6 genetic background and these animals are suitable  
5 for injection or implantation of such tumor cells, as well as other tumor cells described in literature that are immunocompatent for C57BL/6 mice. Thus, the transgenic animals can then be used, for example, to monitor, *in vivo*, tumor progression (*e.g.*, growth) and the efficacy of therapies on tumor regression. For example, where the transgenic animal is tumor-susceptible, it is monitored for expression of a reporter, *e.g.*, luciferase, which is  
10 indicative of tumorigenesis and/or angiogenesis. The monitoring of expression of luciferase reporter expression cassettes using non-invasive whole animal imaging has been described (Contag, C. et al, U.S. Patent No. 5,650,135, July 22, 1997; Contag, P., et al, *Nature Medicine* 4(2):245-247, 1998; Contag, C., et al, *OSA TOPS on Biomedical Optical Spectroscopy and Diagnostics* 3:220-224, 1996; Contag, C.H., et al,  
15 *Photochemistry and Photobiology* 66(4):523-531, 1997; Contag, C.H., et al, *Molecular Microbiology* 18(4):593-603, 1995). Such imaging typically uses at least one photo detector device element, for example, a charge-coupled device (CCD) camera.

The transgenic animals described herein can also be used to determine the effect of an analyte (*e.g.*, therapy), for example on tumor progression where the promoter  
20 induces light generating protein (*e.g.*, luciferase) expression when a tumor develops. Methods of administration of the analyte include, but are not limited to, injection (subcutaneously, epidermally, intradermally), intramucosal (such as nasal, rectal and vaginal), intraperitoneal, intravenous, oral or intramuscular. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal  
25 applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. For example, the analyte of interest can be administered over a range of concentration to determine a dose/response curve. The analyte may be administered to a series of test animals or to a single test animal (given that response to the analyte can be cleared from the transgenic animal).

30 The following examples are intended only to illustrate the present invention and should in no way be construed as limiting the subject invention.

EXAMPLES

**Example 1**  
**Generating the Targeting Cassette and Vector**

5

**A. Creation of the Backbone Vector**

**pTK53:** The 0.5 kb mouse phosphoglycerate kinase 1 promoter was amplified with PGK primers (PGKF, SEQ ID NO:1:

ATCGAATTCTACCGGGTAGGGGAGGCGCTTT; PGKR, SEQ ID NO:2:

10 GGCTGCAGGTCGAAAGGCCCGGAGATGAGG) using mouse genomic DNA (Genome Systems, Inc., St. Louis, MO) as template. This fragment was then double digested with EcoRI and PstI and cloned into the pKS vector (Stratagene, La Jolla, California) which was linearized with the same enzymes. The neomycin gene was amplified with NeoF (SEQ ID NO:3:

15 ACCTGCAGCCAATATGGGATCGGCCATTGAAC) and NeoR (SEQ ID NO:4: GGATCCGCGGCCGCCCCCAGCTGGTTCTTTCCGCCTC) primers using pNTKV1907 (Stratagene) as a template. The 1.1 kb PCR fragment was double digested with PstI and BamHI and cloned into the pKS-PGK vector which was linearized with the same enzymes. This pKS-PGK-Neo vector was used to clone thymidine kinase gene as follows. Primers TKF (SEQ ID NO: 5:

GGATCCTCTAGAGTCGAGCAGTGTGGTTTT) and TKR (SEQ ID NO:6:

GAGCTCCCGTAGTCAGGTTTAGTTCGTCCG) were used to amplify the TK gene

from pNTKV1907 (Stratagene). The amplified 2kb fragment was then digested with BamHI and SacI and cloned into pKS-PGK-Neo vector that was linearized with the same

25 enzymes. This constructed vector was designated as pTK. A synthetic linker F5R5 was made after annealing of two primers (forward primer, F5R51, SEQ ID NO: 7:

GTACATTTAAATCCTGCAGG; reverse primer, F5R52, SEQ ID NO:8:

AGCTCCTGCAGGATTTAAAT). This linker was inserted between Asp718I and

HindIII sites of pTK and the new construct was designated pTK5. A second synthetic

30 linker F3R3 was made by annealing of two primers (forward primer, F3R31, SEQ ID NO:9:

GGCCCCGGGCTTAATTAATGCATCATATGGTACCGTTTAAACGCGGCCGCAAG

CTTGTCGACGGCGCGCCGGCCGGCC; reverse primer, F3R32, SEQ ID NO:10:

GATCGGCCGGCCGGCGCGCCGTCGACAAGCTTGCGGCCGCGTTTAAACGGTACCATATGATGCATTAATTAAGCCCCG). ). This linker was inserted between NotI and BamHI sites of pTK and the new construct was designated pTK53. Schematics of the vectors are shown in Figure 1.

5

#### **B. Introduction of Luciferase**

**pTK-LucYG and pTK-LucR:** The yellow green luciferase gene was isolated from pGL3 vector (Promega) as a HindIII-SalI fragment and was cloned into pGK53 that was linearized with the same enzymes. The new construct was designated pTK-LucYG (8931 bp), shown in Figure 2.

10

The red luciferase gene was isolated from pGL3-red vector (Dr. Christopher Contag, Stanford University, Stanford, Calif.) as a HindIII-SalI fragment and was cloned into pGK53 that was linearized with the same enzymes. The new construct was designated pTK-LucR (8931 bp), shown in Figure 2.

15

### **Example 2**

#### **Insertion of Targeting Sequences**

**A. Generation of vitronectin targeting vector:** The targeting construct pTKLR-Vn was generated by inserting vitronectin (VN) DNA sequences into pTK-LucR vector.

20

Vitronectin (VN) is an abundant glycoprotein present in plasma and the extracellular matrix of most tissues. In a previous study, it was shown that heterozygous mice carrying one normal and one null VN allele and homozygous null mice completely deficient in vitronectin demonstrate normal development, fertility, and survival. This suggests that VN is not essential for cell adhesion and migration during normal mouse development (Zheng, X., et al., Proc Natl Acad Sci U S A 1995 92:12426-30). Mouse vitronectin genomic DNA sequence of 5004 bp was obtained from GenBank database (Accession number X72091). Based on this sequence, a 1.63 kb 3' end vitronectin fragment was amplified (reverse primer, VN1R, SEQ ID NO:11:

25

CTGTATTTAAATCTGCCCCACCCTATTCAGGACAGTAGTC; forward primer, VN1F, SEQ ID NO:12: CCAATGCATCAACCCAGCCAGGAGGAGTGCG) using mouse C57BL/6 genomic DNA as template (Genome Systems, Inc., St. Louis, MO).

30

This fragment was digested with *SwaI* and *NsiI* and cloned into pTK-LucR (linearized with *SwaI* and *SbfI*). This construct was designated as pTK-LucR3. Subsequently, a 2.35 kb 5' end vitronectin fragment was amplified (reverse primer, VN2R, SEQ ID NO:13: AACGCGTCGACTTCGGAGATGTTTCGGGGATAACCAGG; forward primer,

5 VN2F, SEQ ID NO:14:

TTGGCGCGCCCCATAGAGAAGAGACACCAAAGGCACGCTC) using mouse C57BL/6 genomic DNA as template. This fragment was digested with *Sall* and *AscI* and cloned into pTK-LucR vector that was linearized with *Sall* and *AscI*. This construct was designated as pTKLR-Vn. Figure 3 shows the restriction map of pTKLR-Vn vector. The  
 10 polylinker between the neomycin gene and red luciferase gene is used to insert the VEGF promoter or other promoters of interests. The predicted homologous recombination between pTKLR-Vn and vitronectin gene is illustrated in Figure 3A. Upon insertion of the VEGF-LucR transgene cassette, the endogenous vitronectin gene is destroyed. Figure 3B shows the genomic DNA sequence of VN.

15

**B. Generation of Fos targeting vector:** The targeting construct pTKLG-Fos was generated by inserting FosB DNA sequences into pTK-LucYG vector.

FosB is one of the members of the Fos family. It plays a functional role in transcriptional regulation. It has been shown that FosB mice are born at a normal  
 20 frequency, are fertile and present no obvious phenotypic or histologic abnormalities (Gruda et al (1996) *Oncogene* 12:2177-2185). A 28.8 kb genomic region that contains mouse FosB DNA sequence was obtained from GenBank database (Accession number AF093624).

Using this sequence, a 1.71 kb 5' end FosB fragment was amplified (forward  
 25 primer, FosB1F, SEQ ID NO:15: CTGTATTTAAATCCCGTTTCTCACTGTGCCTGTGTC; reverse primer, FosB1R, SEQ ID NO:16: GTCTCCTGCAGGCTTCCTCCTTGTTCCTTGCG) using mouse C57BL/6 genomic DNA as template. This fragment was digested with *SwaI* and *SbfI* and cloned into pTK-LucYG vector that was linearized with *SwaI* and *SbfI*. This construct  
 30 was designated as pTK-LucYG3. Subsequently, a 1.58 kb 3' end FosB fragment was amplified (forward primer, FosB2F, SEQ ID NO:17:

AACGCGTCGACGGATGGGATTGACCCCCAGCCCTC; reverse primer, FosB2R,



SEQ ID NO:18: TTGGCGCGCCCCTTGCCTCCACCTCTCAAATGC) using mouse C57BL/6 genomic DNA as template. This fragment was digested with *Sall* and *AscI* and cloned into pTK-LucYG vector that was linearized with *Sall* and *AscI*. This construct was designated as pTKLG-Fos (Figure 4). The polylinker between the neomycin gene and red luciferase gene is used to insert the VEGFR2 promoter (Example 3, Figure 5A), Tie2 promoter (Example 3, Figure 5B), as well as, other promoters of interests. The predicted homologous recombination between the targeting vector bearing the VEGFR2 promoter (Figure 5A) or the Tie2 promoter (Figure 5B) and FosB gene is also illustrated. As shown in the Figures, the VEGFR2-LucYG transgene cassette and Tie2-LucYG transgene cassette is inserted downstream of FosB gene translational stop signal. Therefore, the targeted transgenic mice should still have a functional FosB gene while expressing the transgenes. Figure 4B shows the DNA sequence of FosB.

### Example 3

#### Insertion of Promoter Sequences of Interest

**A. pTKLR-Vn/VEGF:** Mouse VEGF genomic DNA sequence of 2240 bp that contains a partial VEGF promoter region was obtained from GenBank (accession number: U41383). Accordingly, primers were designed to amplify a 0.69 kb (VF1-VR1A; Table 1) and a 0.98 kb fragment (VF2-VR2; Table 1). It was confirmed that each pair of primers can amplify the predicted product using mouse129SvJ genomic DNA as template.

Table 1

Name	Seq Id No	Sequence
VF1	19	ACCTC ACTCT CCTGT CTCCC CTGAT TCCCA A
VR1A	20	GCTCT GGC GG TCACC CCCAA AAGCA
VF2	21	CCCTT TCCAA GACCC GTGCC ATTTG AGC
VR2	22	ACTTT GCCCC TGTC CTCTC TCTGT TCGC
KF1	23	GCTGC GTCCA GATTT GCTCT CAGAT GCG
KR1	24	TTCTC AGGCA CAGAC TCCTT CTCCG TCCCT
KF2	25	CAGAT GGACG AGAAA ACAGT AGAGG CGTTG GC
KR2	26	GAGGA CTCAG GGCAG AAAGA GAGCG
TF3	27	AGCTT AGCCT GCAAG GGTGG TCCTC ATCG
TF2	28	CAAAT GCACC CCAGA GAACA GCTTA GCCTG C
TR1	29	GCTTT CAACA ACTCA CAACT TTGCG ACTTC CCG

Conditions for PCR amplification are shown in Figure 6. These primers were used for PCR screening of mouse 129/SvJ genomic DNA BAC (bacterial artificial chromosome) library (Genome Systems, Inc., St. Louis, MO). The library, on average, contained inserts of 120 kb with sizes ranging between 50 kb to 240 kb. A large genomic DNA fragment that contained VEGF promoter region was obtained. Southern blot analysis was performed to map the VEGF promoter region. A unique HindIII restriction site was mapped approximately 7.8 kb upstream of the ATG translational start codon of the VEGF gene. The sequences between HindIII and ATG translational start codon are inserted into the polylinker of pTKLR-Vn vector to finish the construction of targeting vector that contains VEGF-LucR transgene (Figure 3A).

#### 15 B. pTKLG-Fos/VEGFR2

A mouse VEGFR2 genomic DNA sequence of 1079 bp which contains a partial VEGFR2 promoter region was published previously (Ronicke, et al., (1996) *Cir. Res.* 79:277-285). Accordingly, primers were designed to amplify a 0.45 kb (KF1-KR1A; Table 1) and a 0.58 kb fragment (KF2-KR2; Table 1). It was confirmed that each pair of primers can amplify the predicted product using mouse 129SvJ genomic DNA as

template. DNA sequences for these primers are shown in Table 1 above and PCR amplification conditions are shown in Figure 6. These primers were used for PCR screening of mouse 129/SvJ genomic DNA BAC library. From the screening, a BAC clone that contained a large genomic DNA fragment was obtained. Based on the published restriction map of VEGFR2, a 4.5 kb HindIII-XbaI fragment that covers the VEGFR2 promoter region was subcloned from the VEGFR2 BAC clone into the pBluescriptSK vector (Stratagene, La Jolla, CA) that was linearized with HindIII and XbaI. The VEGFR2 promoter sequences of 4.1 kb, spanning from a HindIII site to the ATG translational start codon, are subcloned into the polylinker of pTKLG-Fos vector to construct the targeting vector that contains VEGFR2-LucYG transgene (Figure 5A).

### C. pTKLG-Fos/Tie2

Mouse Tie2 genomic DNA sequence of 477 bp which contain a partial Tie2 promoter region has been published previously (Fadel et al (1998) *Biochem. J.* 330(Pt. 1):335-343). Accordingly, primers were designed to amplify a 0.45 kb (TF3-TR1; Table 1) and a 0.47 kb fragment (TF2-TR1; Table 1). It was confirmed that each pair of primers amplified the predicted product using mouse 129SvJ genomic DNA as template. DNA sequences for these primers are shown in Table 1 above. PCR amplification conditions are shown in Figure 6. These primers are used for PCR screening of mouse 129SvJ genomic DNA BAC library. From the screening, a BAC clone containing a large genomic DNA fragment of the Tie2 promoter region was obtained. Based on the published Tie2 genomic DNA restriction map (see, Dumont et al., *supra*), a 10.5 kb Asp718-EcoRV fragment containing the Tie2 promoter region was subcloned ifrom the Tie2 BAC clone into the pSK vector that was linearized with Asp718 and EcoRV. The Tie2 promoter sequences of about 6.8 kb, spanning from the Asp718 site to the ATG translational start codon is subcloned into the polylinker of pTKLG-Fos vector to construct the Tie2-LucYG targeting vector (Figure 5B).

**Example 4****Generation of Transgenic Mice Carrying the Constructs of the Present  
Invention**

5       **A. General Procedure:** Figure 7 depicts a generalized description of generation of transgenic mice using the targeted transgenic vectors described in Example 3. Details regarding embryonic stem (ES) cell culture, transfection, blastocyst injection and implantation to a pseudopregnant foster are described, for example, in Hogan et al (1994) "Manipulating the Mouse Embryo, A Laboratory Manual. Second Edition", Cold Spring  
10 Harbour Laboratory Press.

After construction the targeted transgenic construct are transfected into C57BL/6 embryonic stem (ES) cells. (Genome System Inc., Genome Systems, Inc., St. Louis, MO) through electroporation. The antibiotic G418 is used to select for cells in which the DNA construct containing the Neo gene is integrated, either randomly or by homologous  
15 recombination. The nucleoside analog gancyclovir is converted by TK to a cytotoxic derivative. DNA that has integrated by homologous recombination lose the TK gene and are resistant to the drug, whereas cells that have incorporated the DNA randomly are likely to retain the TK gene. Thus, cells containing random integrations into a chromosomal location that allows the expression of the TK gene are killed. The G418  
20 and gancyclovir resistant clones are then be screened by PCR and Southern blot analysis and those that have homologous DNA recombination is used for FVB/N blastocyst injection (Genome System, Inc.). Between 4-16 blastocysts are transferred to the uterus of a pseudopregnant foster mother. The pups are typically born 17 days after the transfer. Either random bred mice or F1 hybrid mice make suitable recipients. Females of certain  
25 random-bred stocks (*e.g.*, CD1 mice, from Charles River Laboratories) have very large ampullae, which makes oviduct transfer easier. These mice also generally make good mothers. Alternatively, F1 hybrid females (*e.g.*, B6 x CBA F1) can be used as recipients. Although their ampullae are smaller, make exceptionally good mothers, rearing litters as small as two pups. See, for example, Hogan et al. (1994), *supra*.

### B. Screening for homologous DNA recombination positive ES cells

1) **pTKLG-Fos/VEGFR2:** Analysis of homologous DNA recombination between pTKLG-Fos/VEGFR2 targeting vector and the FosB gene is carried out using Southern blot analysis as shown in Figure 8. Genomic DNA prepared from G418 resistant ES cells is digested with PvuII and probed with probe A to confirm the 5' end DNA recombination. PvuII digestion of DNA bearing homologous recombination reveals two separate bands of 8.2 and 4.0 kb, whereas digestion of DNA from homologous recombination negative clones reveals only the 8.2 kb band. The 3' end of DNA recombination is tested by hybridizing NotI digested DNA with probe B. NotI digestion of DNA bearing homologous recombination will reveal two separate bands of >8.2kb and 5.0 kb, whereas digestion of DNA from homologous recombination negative clones will only reveal the >8.2 kb band. Once homologous DNA recombination is confirmed, positive clones are selected for FVB/N blastocyst injection.

2) **pTKLG-Fos/Tie2:** Analysis of homologous DNA recombination between pTKLG-Fos/Tie2 targeting vector and the FosB gene is analyzed by Southern blot in a similar manner as described above for pTKLG-Fos/VEGFR2. Once homologous DNA recombination is confirmed, positive clones are selected for FVB/N blastocyst injection.

3) **PTKLR-Vn/VEGF:** Analysis of homologous DNA recombination between pTKLR-Vn/VEGF targeting vector and the vitronectin gene is analyzed by PCR. DNA primers designed according to the predicted homologous recombination, are listed in Table 2.

**Table 2**  
PCR primers for analysis of homologous DNA recombination between pTKLR-Vn/VEGF targeting vector and the vitronectin gene

5

<b>5'end primers</b>	
<b>F51</b> 5'- CCCAGTGTCTCTGATTAGGGAGAGCACCTGAG -3'	SEQ ID NO:30
<b>R51</b> 5'- CCAGACTGCCTTGGGAAAAGCGCCTC -3'	SEQ ID NO:31
<b>F52</b> 5'- CAGTGAGAGTCTTCTCTGTCCCTCAATCGGTTCTG -3'	SEQ ID NO:32
<b>R52</b> 5'- TGGATGTGGAATGTGTGCGAGGCCAG -3'	SEQ ID NO:33
<b>3'end primers</b>	
<b>F31</b> 5'- AATCAAAGAGGCGAACTGTGTGTGAGAGGTCC -3'	SEQ ID NO:34
<b>R31</b> 5'- CGGCTCCCCAAAATGTGGAAGCAAGC -3'	SEQ ID NO:35
<b>F32</b> 5'- GAATCCATCTTGCTCCAACACCCCAACATC -3'	SEQ ID NO:36
<b>R32</b> 5'- CGCCTCCTCTCCCCAGTCTCCCCTTG -3'	SEQ ID NO:37

Primers F51-R51 and F52-R52 amplify a 1799 bp and a 1841 bp DNA fragment respectively from the 5' end of the transgene that is integrated into the vitronectin site through homologous DNA recombination, whereas primers F31-R31 and F32-R32  
10 amplify a 3549 bp and a 3428 bp DNA fragment respectively from the 3' end of the transgene that is integrated into the vitronectin site through homologous DNA recombination. Clones that allow successful amplification of both the 5' end and 3' end of the integrated transgene are selected for FVB/N blastocyst injection.

15

### C. Analysis of chimeric mice

The pups developed from injected blastocysts contain chimeras, as can be identified by their agouti coat color when an ES cell derived from a mouse having a dark coat color (e.g., C57BL/6) is injected into the blastocyst of a light coat color animal (e.g., FVB/N, genotype B/B). DNA analysis (e.g., Southern blotting, PCR) is conducted to  
20 further confirm the presence of the transgene in these pups as described above in Section B. These animals may be obtained commercially, for example from The Jackson Laboratory, Bar Harbor, MN.

#### **D. Generating targeted transgenic C57BL/6 mice with white coat color**

Breeding of the chimeric mice generates homozygous targeted transgenic mice, as depicted in Figure 9. The targeted mice are used to monitor gene expression through the measurement of luciferase mediated light emission from the mice. In a preferred embodiment, the targeted mouse has a light coat color (e.g., white coat color), because the black colored coat (an example of a dark coat color) of C57BL/6 mice can absorb light emitted from the body and may interfere the sensitivity of the bioluminescence assay. An inbred mouse strain C57BL/6-Tyr C2j/+ strain (Jackson Laboratory, Bar Harbor, MN) is available for this purpose. This strain of mice have white color coat, yet they still have the same genetic background as C57BL/6 mice except that the gene responsible for the black coat color is mutated. Unfortunately, C57BL/6-Tyr C2j/+ ES cells are not currently available. Therefore, the designed breeding program illustrated in Figure 9 is aimed to generate mice that are homozygous for the target transgene and have white coat color. C57BL/6 ES cells are prepared as described above and introduced into a suitable blastocyst (e.g., from the FVB/N strain of mice). The blastocysts are implanted into a foster mother. Chimeric mice are shown in Figure 9 as white animals with black and green patches. Chimeric animals are bred with C57BL/6-Tyr C2j/+ mice to create F1 hybrids. Subsequent breeding of the F1 hybrids generates several type of mice, including the one that is homozygous for the target transgene and has a white coat color (shown in Figure 9 as b/b; L/L), which is used for *in vivo* gene regulation monitoring.

A C57BL/6 mouse and a C57BL/6-Tyr C2j/+ mouse are considered to be substantially isogenic. Accordingly, the method of the present invention exemplified in Figure 9 provides a means for generating breeding groups of substantially isogenic mice in a selected genetic background carrying at least one transgene of interest.

#### **E. Dual luciferase targeted transgenic mice**

As described above, two targeting vectors are generated. PTKLR-Vn carries a red luciferase gene and is targeted into vitronectin locus. PTKLG-Fos carries a yellow-green luciferase gene and is targeted into FosB locus. A number of promoters, including VEGF promoter, VEGFR2 promoter, and Tie2 promoter are cloned into these vectors, as described above. Subsequently three type of targeted transgenic mice are generated.

VEGF mice carry VEGF promoter-red luciferase transgene (VEGF-LucR) integrated into vitronectin locus. VEGFR2 mice carry VEGFR2 promoter-yellow-green luciferase (VEGFR2-LucYG) transgene integrated into FosB locus. Tie2 mice carry Tie2 promoter-yellow-green luciferase (Tie2-LucYG) transgene integrated into FosB locus. Through a  
5 breeding program illustrated in Figure 10, dual luciferase targeted transgenic mice are produced, carrying both of the VEGF-LucR and the VEGFR2-LucYG transgenes. The degradation of luciferin by yellow-green luciferase and red luciferase generates lights that emit at 540 nM and 610 nM respectively. These wavelengths of light are measured individually using a photo-counting camera (intensified CCD). Therefore, both VEGF  
10 expression and VEGFR2 expression, for example, can then be monitored in the same mouse at the same time.

As is apparent to one of skill in the art, various modification and variations of the above embodiments can be made without departing from the spirit and scope of this  
15 invention. These modifications and variations are within the scope of this invention.



CLAIMS

What is claimed is:

1. A vector for use in generating a transgenic non-human mammal, said mammal having at least one single-copy, non-essential gene in its genome, said vector comprising
  - (a) a first selectable marker and a reporter expression cassette, said reporter expression cassette comprising a transcriptional promoter element operably linked to a light-generating protein coding sequence, and
  - (b) targeting polynucleotide sequences homologous to a single-copy, non-essential gene in said mammal's genome, said targeting polynucleotide sequences flanking (a), wherein (i) the length of the targeting polynucleotide sequences are sufficient to facilitate homologous recombination between the vector and the single-copy, non-essential gene, and (ii) said transcriptional promoter element is heterologous to the single-copy, non-essential gene.
2. The vector of claim 1, wherein said first selectable marker provides a positive selection.
3. The vector of claim 2, wherein said first selectable marker is selected from the group consisting of neomycin phosphotransferase II, xanthine-guanine phosphoribosyltransferase, hygromycin-B-phosphotransferase, chloramphenicol acetyltransferase, and adeninephosphoribosyl transferase.
4. The vector of claim 3, wherein the first selectable marker is neomycin phosphotransferase II.
5. The vector of any of claims 1-4, further comprising a second selectable marker, wherein at least one target polynucleotide sequence is located between said second selectable marker and (a).
6. The vector of claim 5, wherein said second selectable marker provides a negative selection.

7. The vector of claim 6, wherein said second selectable marker is selected from the group consisting of adenosine deaminase, thymidine kinase, and dihydrofolate reductase.

8. The vector of any of claims 1-7, wherein the transcriptional promoter element is selected from the group consisting of an inducible promoter, a repressible promoter, and a constitutive promoter.

9. The vector of claim 8, wherein the transcriptional promoter element is selected from the group consisting of VEGF, VEGFR, and TIE2.

10. The vector of any of claims 1-9, wherein the sequences encoding the light-generating protein are obtained from either procaryotic or eucaryotic sources.

11. The vector of claim 10, wherein the light generating protein is a luciferase.

12. The vector of any of claims 1-11, wherein said reporter expression cassette further comprises other control elements.

13. The vector of claim 12, wherein said control elements are selected from the group consisting of transcription enhancer elements, transcription termination signals, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

14. The vector of any of claims 1-11, wherein said vector is circular.

15. The vector of claim 14, wherein said vector contains at least one restriction site whose cleavage results in a linear vector having the following arrangement of elements: targeting polynucleotide sequence - (a) - targeting polynucleotide sequences.

16. The vector of claim 5, wherein said vector is circular.

17. The vector of claim 16, wherein said vector contains at least one restriction site whose cleavage results in a linear vector having the following arrangement of elements: target polynucleotide sequence - (a) - targeting polynucleotide sequences - (second selectable marker).

18. The vector of any of claims 1-17, wherein the coding sequences of the reporter expression cassette comprise codons that are optimal for expression in a host system into which the expression cassette is to be introduced.

19. The vector of claim 18, wherein said mammal is a rodent.

20. The vector of claim 19, wherein said mammal is a mouse.

21. The vector of any of claims 1-20, wherein said targeting polynucleotide sequences from single-copy, non-essential genes are selected from the group consisting of vitronectin, *fosB*, and galactin 3.

22. A method of producing a transgenic, non-human mammal, said mammal having at least one single-copy, non-essential gene in its genome, comprising  
transfecting an embryonic stem cell of said mammal with a linear vector comprising

(a) a first selectable marker and a reporter expression cassette, said reporter expression cassette comprising a transcriptional promoter element operably linked to a light-generating protein coding sequence, and

(b) targeting polynucleotide sequences homologous to a single-copy, non-essential gene in said mammal's genome, said targeting polynucleotide sequences flanking (a), wherein (i) the length of the polynucleotide sequences are sufficient to facilitate homologous recombination between the vector and the single-copy, non-essential gene, and (ii) said transcriptional promoter element is heterologous to the single-copy, non-essential gene;

selecting embryonic stem cells which each have said first selectable marker and reporter expression cassette integrated into its genome;  
injecting said embryonic stem cells into a host embryo,  
implanting said embryo in a foster mother,  
maintaining said foster mother under conditions which allow production of an offspring that is a transgenic, non-human mammal carrying said reporter expression cassette.

23. The method of claim 22, wherein said offspring is capable of germline transmission of said reporter expression cassette.

24. The method of claim 23 further comprising breeding said offspring with a mammal, wherein the mammal is substantially isogenic with the embryonic stem cells, wherein said breeding yields transgenic F1 offspring carrying said reporter cassette.

25. The method of claim 24, further comprising breeding a first F1 offspring carrying said reporter cassette with a second F1 offspring carrying said reporter cassette, wherein said breeding yields transgenic F2 offspring carrying said reporter cassette.

26. The method of any of claims 22-25, wherein said mammal is a mouse.

27. The method of claim 26, wherein said embryonic stem cells are derived from a mouse having a dark coat color.

28. The method of claim 27, wherein said mammal substantially isogenic with the embryonic stem cells has a light coat color.

29. The method of claim 28, wherein said F2 offspring carrying said reporter cassette has a light coat color.

30. The method of claim 29, wherein said embryonic stem cells are derived from a C57BL/6 mouse having a dark coat color, and said mammal substantially

isogenic with the embryonic stem cells is a C57BL/6-Tyr C2j/+ mouse having a light coat color.

31. The method of any of claims 22-30, wherein the light generating protein is a luciferase.

32. A transgenic, non-human mammal, comprising at least one single-copy, non-essential gene in its genome, wherein (i) at least a portion of at least one single-copy, non-essential gene is replaced by polynucleotide sequences heterologous to the gene, such that the gene cannot produce a functional gene product, and (ii) said polynucleotide sequences comprise a first expression cassette which has been introduced into said mammal or an ancestor of said mammal, at an embryonic stage, said first expression cassette comprising

a first selectable marker,

a first transcriptional promoter element heterologous to the gene, and

light-generating protein coding sequences, wherein said light-generating protein coding sequences are operably linked to said promoter element.

33. The transgenic, non-human mammal of claim 32, wherein the single-copy, non-essential gene is selected from the group consisting of vitronectin, *fosB*, and galactin 3.

34. The transgenic, non-human mammal of any of claims 32-33, wherein the first selectable marker is selected from the group consisting of neomycin phosphotransferase II, xanthine-guanine phosphoribosyltransferase, hygromycin-B-phosphotransferase, chloramphenicol acetyltransferase, and adeninephosphoribosyl transferase.

35. The transgenic, non-human mammal of any of claims 32-34, wherein the first transcriptional promoter element is selected from the group consisting of an inducible promoter, a repressible promoter, and a constitutive promoter.

36. The transgenic, non-human mammal of claim 35, wherein the first transcriptional promoter element is selected from the group consisting of VEGF, VEGFR, and TIE2.

37. The transgenic, non-human mammal of any of claims 32-36, wherein said expression cassette further comprises other control elements.

38. The transgenic, non-human mammal of claim 37, wherein said control elements are selected from the group consisting of transcription enhancer elements, transcription termination signals, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

39. The transgenic, non-human mammal of any of claims 32-38, wherein said non-human mammal comprises a second single-copy, non-essential gene in its genome, wherein (i) at least a portion of the second single-copy, non-essential gene is replaced by polynucleotide sequences heterologous to the second gene, such that the second gene cannot produce a functional gene product, and (ii) said polynucleotide sequences comprise a second expression cassette which has been introduced into said mammal or an ancestor of said mammal, at an embryonic stage, said first expression cassette comprising

a second selectable marker,

a second transcriptional promoter element heterologous to the second gene, and light-generating protein coding sequences, wherein said light-generating protein coding sequences are operably linked to said second promoter element.

40. The transgenic, non-human mammal of claim 39, wherein the first transcriptional promoter element in the first expression cassette is different from the second transcriptional promoter element in the second expression cassette.

41. The transgenic, non-human mammal of claim 40, wherein the light-generating protein in the first expression cassette can produce a different color of light relative to the light-generating protein in the second expression cassette.

42. The transgenic, non-human mammal of any of claims 32-41, said mammal being a rodent.

43. The transgenic, non-human mammal of 42, said rodent being a mouse.

44. The transgenic, non-human mammal of any of claims 32-43, wherein the sequences encoding the light-generating protein are obtained from either procaryotic or eucaryotic sources.

45. The transgenic, non-human mammal of claim 44 wherein the light-generating protein is a luciferase.

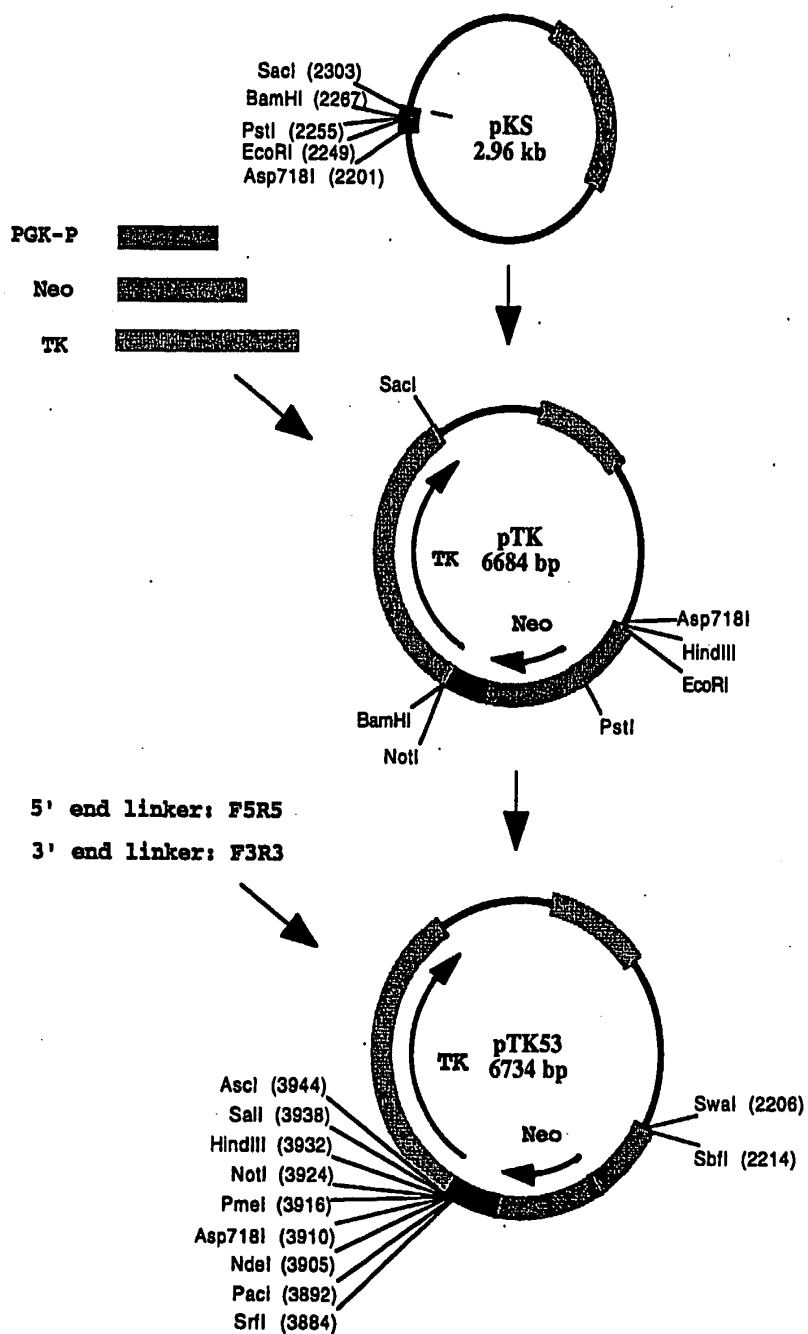


FIG. 1



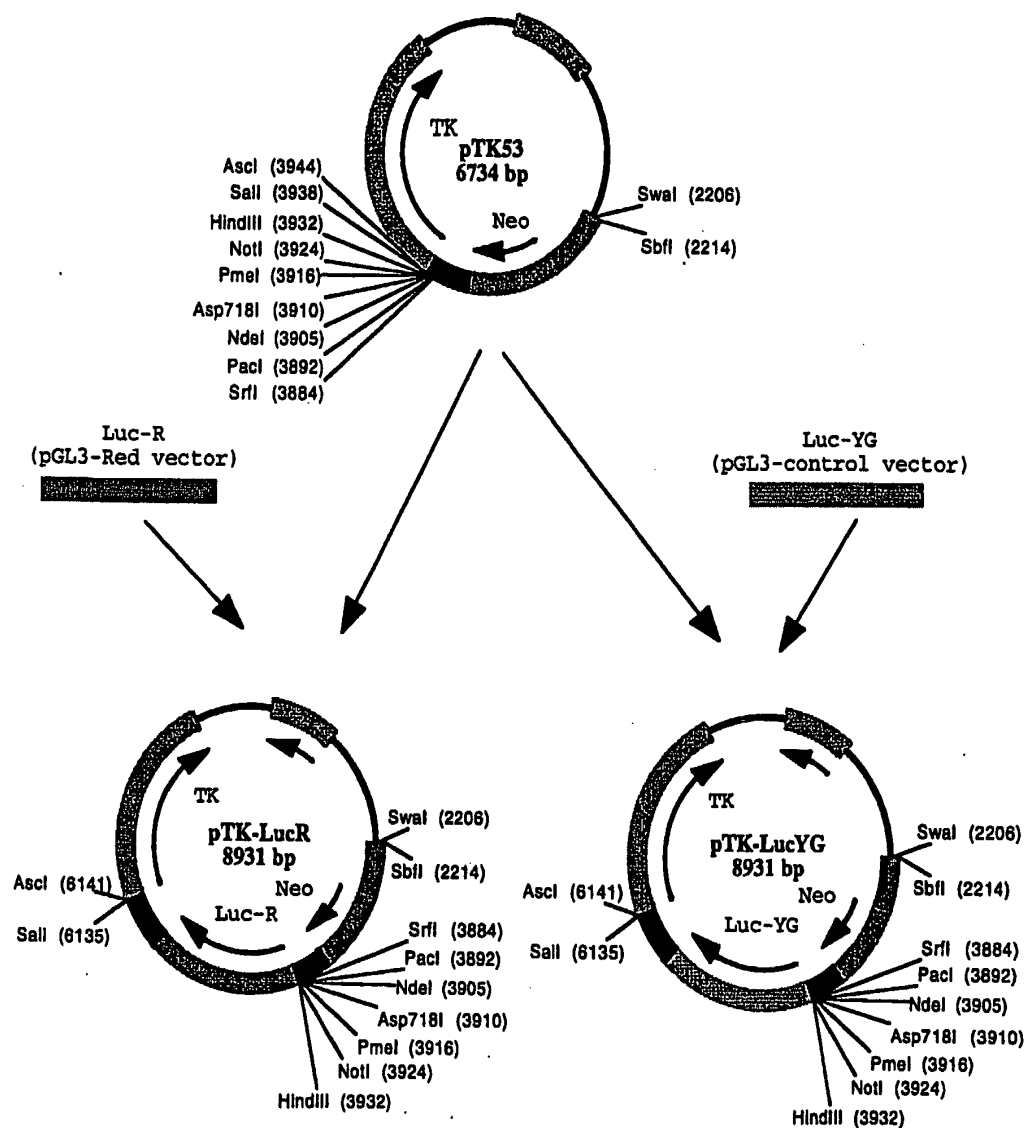


FIG. 2

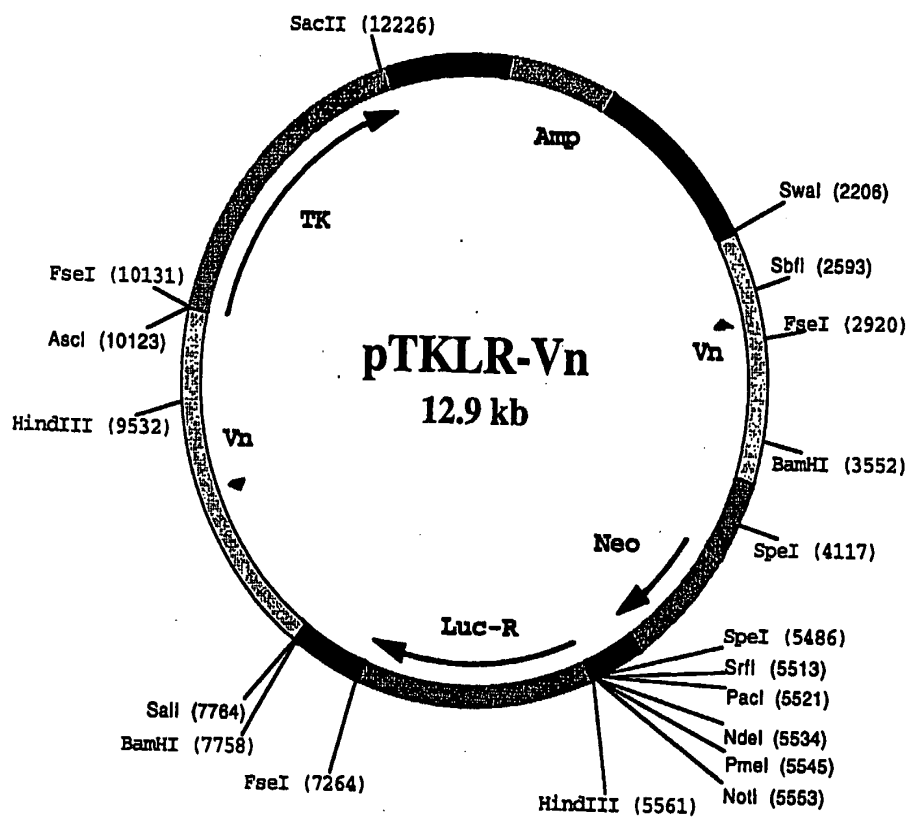


FIG. 3A

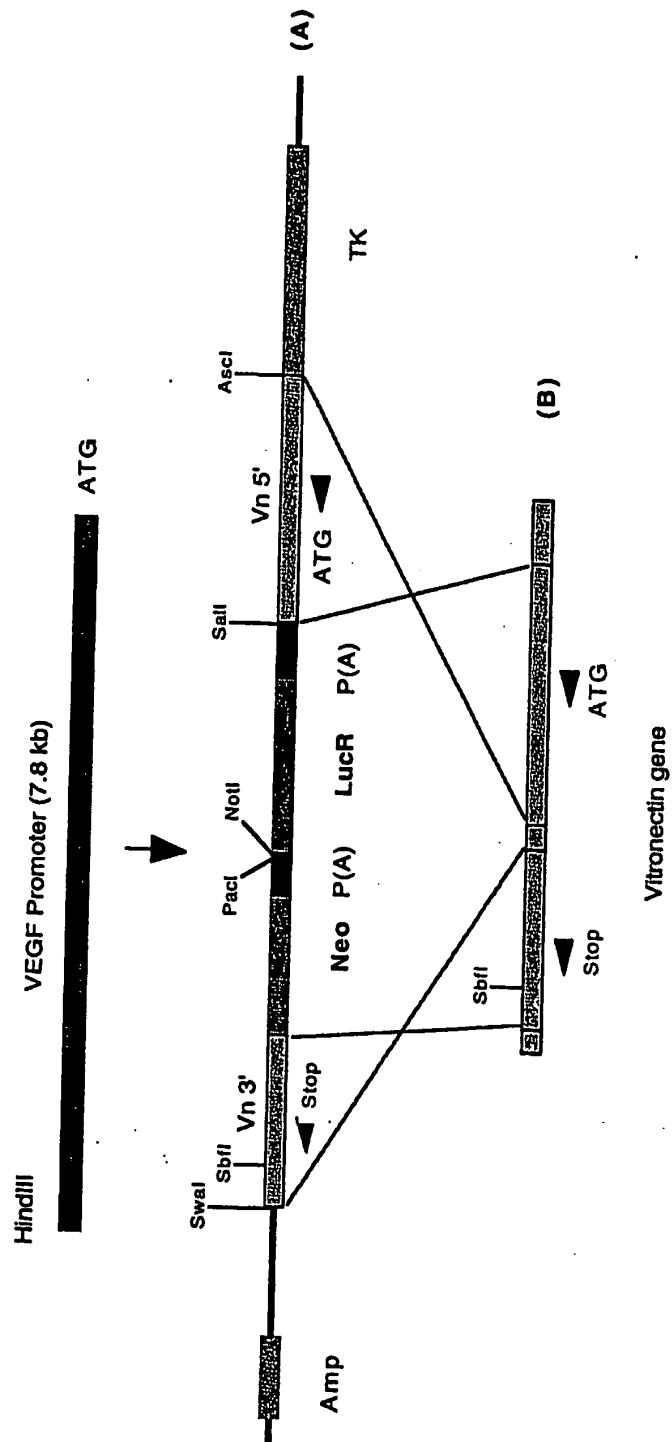


FIG. 3B

1 TCCAGCACC TGTTCCTAC GTCCCGCCG TCTCAGTGA ACTTCATGT TAAAGAGGC TCACCCCGG AGGTTGTGT GCCACAGAG GAGGGTCTT  
AGTGGGTGG ACAGAGGTG CAGGGGCGG AAGATCAAT TGAAGTACA ATTCTCTCG ATGCGCGCC TCCACACCA CGGTGTCTC CTTCACAGA

101 CCCACAGCC CCCAGTGTCT CTGATTAGG GAGGACACT GAGCCCACT AGAGTCTCT CTGTCCCTCA ATCGGTCTG AAATTCGCA CTGCGCTGC  
GGGTGTTCG GGTTCACAG GACTAAATCC CTCTCTGGA CTGGGTGAC TCTCAGAGA GACAGGGAT TAGCCAGAG TTTAGGGGT GAGCGGAGG

201 TTATCCAGG GACAGGCTG CCCACCTAT TCAGGACAT AGTCTTAAC TCGTAGCCA CAGACTTTT ATGCGCTGG GAGAAAGAG TAGGCTCTT  
AATAGTCCC CTGTCCGAC GGTGGGATA AGTCTGTCA TCAGATTTC AGCATCGGT GTCTGAAAA TACCCGACC CTCTTCTCT ACTCGAGGA

301 GAGGCTAGC CGAGTGGCT CTGATTCCTA CTCTCAGAG GTCCCGCAG CCAGCCATA CTGAGCANT GAGGTTGGT AGGAGGAT CACAGATCC  
CTTCGATCG GCTCACCGA GACTAAGAT GAGAGTCTC GAGCCGTGG GTTCGGTAT GACTGTGTAC CTCCACCCA TCCCTCTTA GTGTCTAGG

401 ACTCGCGGG TTCTAGGTT GACTCGGTG TATTTCTCT AAGAAAGAA TGGAAAAAG GTATGTGAG ATTCTCCCT ATCTCTCCA CTGTCCCAA  
TGAAGCGCC AAGATCCAA CTAGCCATC ATAAACAGAC TTCTTTCTT AACTTTTTTC CATACACTC TAGACGGAC TAGACAGGT GAGCGGGTT

501 GAGGATAAA GCTTTTCTT CAGAGGCAA AGTGAACAT CACCAAGAG ATATGTCTC CATCTACAG CTGTGTTCAG CACCCAGGA CCAGAGCTG  
CTTCTATTT CGAAGAAAG GTCTTCTCT TCACTGTAG GTGTTCCTC TATTACATG GTAGATGTC GACACAGTC GTGGTCTCT GTTCTGTAC

601 CAGGCAAGC CTAGCCAAA CCAGTCTAG GAGTAGAAG GGGTCCAC CTCAGAGAA GAAATAGAG CTCTGATGG GCTCCAGGT GCGAGTACA  
GTCTGTTCG GTCAGTTCT GTCTAGATC CTGAGGTG GAGGTCTCT CTCTATCTC GAGACTTAC GAGACTTAC GAGACTTAC GAGACTTAC

701 AGCCAGTCA TATCATATC ATAGTTGTG TAGGTTCTA GCGGCTCTC CTGCTGGAG AACAAAGAG ACCAGATTG AGGTGTAA GAGCGGAT  
TGGTCAAGT ATATATATG TATCAACAC ATCCAGGAT GGGTGGAG GAGCGACTC TTCTTTCTT TGTCTAAT TCCACTACT GCTGCTCA

801 TGAAGTCTG GCTGGTCTG TGGCCAGCC CTGCGGTGA AGATAGAGC TTTCGGCTC TAGGCTTGA CTCTGTCTT TTGGTGGG CAGATGGGA  
AGCTGAGAC CGAGCCGAC ACCGTTGGG GAGCGGACT TGTATGCGG AAGCGGAG ATCGGATCT GAGAGAAA AAGCGAGCC GTCTCAAGT

901 TAGGAGCCA GTAGCTTGA TCGCGCGCC CATAGCAGG TCCACTTTC CTGCGACAC ATCGGATTC GCGGTATGA ATTGGGTTG TCTGCTCCA  
ATTCTCGGT CACTGTACT TCGCGCGCC GTATGTTCG AGGTGAAGG GAGCGTGTG TAGGCTTGA GCGGCTACT TACCGTACC GAGATTTGT

1001 TCTGTAAGC GGAAGGGTT AATCACTTG GAGATCTCT GCTTTGATT CTCCAGCAA GTGTCTCTC TATCTATTA TCTATCTTA TCTATGATC  
AGCATATTC GCTTCCCAA TTAGGTGAC GGTCTAGAC GGAATCTAA GAGGTCTCT CACAGACAG ATAGATAAT AGATAGAA TATATCTAG

1101 TATCTATTA TCTATGATC TATCTATTA TCTATGATC ACTCTATC CTATCTATG ATCTATCTAT CTATCTATC CTTACTACT TATCTATTA  
ATAGATATG AGATATGAG ATAGATATG ATAGATGGA TGTATGATG GATATGATC TAGATAGAT TAGATAGAT GATATAGAT GATATAGAT

1201 CTTATTTAT TTTTCTTGA AACAGATCT TAGCAGTAC CTATGCTGG TTTCGACTC ACTATGAGC CATCTATGC CTCTACTC  
GATATATTA ACAGAAACT TTGTCTTGA ATCTGATGA ATCTGATG GATACGACT AAGCTTGA TGTACTTGG TATCTATGC GAGATTTAG

1301 ACAAGATCC ACTTCTCTG GTCTCTGAT GCTGGATTA AAGCATGTC CCATCTACC CAGCTCCAT AGGACCTTA GACACATTT GCTATGCTT  
TGTCTTAGG TGAAGGACA CAGAGCTCA CGACCTAAT TTCTGTAC GGTGATGTC GTGAGGTCA TCTTGAAAT CTGTGTAA GATACGAA

1401 GCTTAGACA CACACTCAG TCGCGCGCC CGACCTTCC TGTCTAGAG TTCTTCCAT GCTCTCTCA CTGTATGCT TGAATCTG CCGCATCCA  
CGATTTCTT GTTGTAGTC AGGGTCCGG GGTGGAGGG ACAGATCTG AAAAAAGGTA GAGAGAGGT GACATAGGA ACTTAGAGC GGGTAGGCT

1501 AACCCCTCAG CCGCGTCCG CTCTCTTCC TGTGTAGCG AAGTCCAG GTATGGGAT CAATAGAGC CAGGCTCAT CCGCCAAAG TCAAGAGAG  
TTGGGATTC GCGGTCCGG GAGGAGAGG ACAGATCCG TTTCAGGTT CATTAGGAT GTTATCTCG GTTCGAGTA GGTGGTCTT ACTTGTCTC

1601 CAAAGTCTG CAGAGCAA CAGCTCTGA TGTATGTT CAGATTOCA GCGGCTCC CTGGAAGCC CCATCTAC AGCCGATTT CAGAGAAAG  
GTTCAGATC GTTCTGTT GTCTGAGAT AGCTACACA GTGTCAAGT CCGGGAGGG GAGCTTCCG GGTGATGTC TCGGTCAA GGTCTCTTC

1701 AAGCCAGCT TCTCTCTCT CCAATCCAG GATCTCTCC CAGAGAGGA GTTGGAAAT GTTCTCCAG CTGTCCGCT GAGCAGGC AAAGTCTCA  
ATCCAGCCA TGAATCTT CCAATCCAG GATCTCTCC CAGAGAGGA GTTGGAAAT GTTCTCCAG CTGTCCGCT GAGCAGGC AAAGTCTCA

1801 AACAGGCTG CAGAGAGCT GCTTCCGAC TCTCTCTCC TGGGTCTCT GTGAAATTC TACTCCGAT ACTCTCTCC TGAAGAGG AACAGTCC  
TTGTCCGAC GTCTCTCTA CGAGAGCTG AGGAGAGG ACGAGAGG CTGAAATTC TACTCCGAT ACTCTCTCC TGAAGAGG AACAGTCC

1901 ATCAGAGAG ATCTGACAA GCGAGAGAG AATCATGGA TAGACAGG ACTCCACAC CTGCGGCTT CTCTCCAC CTGATGACC TTGAGAGT  
TAGTCTCTC TAGATGTT CTCTCTCTC TTATGACTT ATCTTGTCC TGAAGTGTG GAGCGGCAA GAGGAGTGG GACTATGAG AACTCTCTA

2001 AGACCTTTC CCGGCTCTG TAGGTTGG CAGGAGGCG GAGCCTTCA TCAATCTT CTGTGTGCT ACTGAGCT TCGAGATG TTGCGGAT  
TCTGGAAAG GCGGCTGAC ATTCCAGCC GTCTTCCCG CTGCGAGCT AGTTGTACA GAGCAGAG GAGCAGAG GAGCAGAG GAGCAGAG

2101 AACAGGCTG AGGAGGCTT CCAATCCAG CCAATCTCA CTACCTGAA AGACAGAT CAGAGAGGT AGGAGTACC GCTGCGACA GAGAGTCC  
TGTCCAGG TCTGCGGTA GAGGTTTCC GGTCTGACT CTACCTGAA AGACAGAT CAGAGAGGT AGGAGTACC GCTGCGACA GAGAGTCC

2201 TATATCTTA ACTGCTCTC AGCTCTCTC GAGTCCCTG ACTCTCTT CTTCAGACT CCGGCGAG TCGATGAC CTTTGGCT GCTCTGACT  
ATATGATTT TGAAGAGAG TGAAGAGAG CTTGAGGAG TGAAGAGAG GAGTGTGA GAGGCTCTC TGAAGAGAG GAGGCTCTC TGAAGAGAG

2301 TAGCTCTCT ACTTGAACA AGTATGCTT CCGCTGAGC TTATGAGAG TGAAGGAGC ATCGATGCG CCGTCAATC CCGAGATC TTGATATTT  
ATCCAGCCA TGAATCTT TCAATCAAA GGGGACTTC AACTGCTCT TGAAGGAGC ATCGATGCG CCGTCAATC CCGAGATC TTGATATTT

2401 TTGGGTGAG CAGGCTCTC TCGCTCTCA TCTAGCTAT AGCAGTACT CCGTGAACA GGGGAACTG TGTGAGAGC AGATGAGCT AAGGAGATC  
AACCCATGG TCGCGGATG ACCGAGAT AGATGAGTA TGTCTATG GGGGAACTG TGTGAGAGC AGATGAGCT AAGGAGATC

2501 CAGCGGAC CAGACTCTC CATAGATCA CTTCCGAGC CAAAGAGGA CCAATCTT AGATGAGTA AGGCTCAA GGGCTTCCA CTGCAAGTT  
GCTGCGGTC GTCTGAGAG GTATCTCAT GAGGCTTCC GTTCTCTCT GGTGAGAA TCTAGGACT TCGGAGTT CCGGAAAGT GAGGTCTCA

2601 CTCTCTCTG AAATCTCTG GTCCTTCTT CAGTGTCTC GGGGCTTAG ATCTCTCT GTTCTCTC TTAGGGCT GGGGCTCTG GCTTCTCTC  
GAGGAGAGC TTTGATGCT GTCTCTCTT CAGGAGTCA CCGGAGTCC TAGAGGAG TGAAGGAGT AAATCTCTA GGGGCTCTG GCTTCTCTC

2701 AGATCTAGC AAGGCTCTG CTTTAGACT GGGTCTCT CAGGATTTA GTTCACCGG TGAAGAGGT TTCTCGGTT GCACAGGTT GTTGTATTT  
TCTTAGATC TTCCAGAGC CGAATCTCA CCGGAGGAG CTTCTAAT CCAATGCGCC AACTCTCTC AAGAGCCAA GGTGTGCA CACATCTAC

2801 TTCTTGGCT CTTCCAGTA GTATAGCT CAAATATAT CTTCTGAT AGTGAAGC TCGCGGCG TACTCTAG CAGAGGCG AGCAGTAT  
CAGAGGAG GAGGCTCAT CATATGAG GTATATGA CAGAGGTA TCACTTCTC AAGGCGCG AATGAGTCC GTCTTCCCG TGTCTACTA

2901 GTGAGCTCT GAGGAGAGC CAGGCGGAG CAGAGGAG TCTGAGTCA CTTTGGGCT TCACTCTC CATGATGTC GCAGGAGC TGTATATTA  
CAGTCCACA CTTCTTCCG GGTCCGAGT GGTGTCCCG AGACTTGGT AGTGAAGC AGTGAAGC GTATGAGC CAGGAGC GTTGTCTG AGACTATCA

FIG. 3C-1

3001 AGTCAGAGC TGTGACACT GACACTTCTT OCTGCCATG AAACCTGAG TGCAGCGCC CTTCATGAC TCTATGGAG GGAATATCG GTTTCAGCC  
TCACTTTTCG AGCATGTGA CTGTGAAGAA CGACCGGTAC TTTCGGACTC AGTCCGCGG GAACGTACTG AGATACCCCT CCTTATATG CAAATGTCCG

3101 CAATCAGGG CACTTCCCA ACTGCACCTT CCTGAGTAC CCACCAATCC OCTCCACAC CTTCGTGAG CAGAGAAAC CATGCCACA GGCCTGATAT  
GTAGATGCC GTGACGGGT TCGACGTGAA GCGATCCATG GGTGTGTAGG GGAAGGTGTG GAACCACTCG GTCTCTTTTG GTACCGTGTG GGCCTGATAT  
CCGATCATA

3201 GAAAAAGGC CTCAGGGGTG CCAATGCCAG OCTCTAGGCC AGGCGCTTGG CAGGCTGGGC GCGGAGCTTC TCGAATCTCG CTGTCTGCC TGA AAAAAGA  
CTTTTTCGG GAGTCCGAC GGTACCGTCC GAGATCCGG TCCCGAACC GTTCGACCG CCGCTCGAG ACCTTAGAGC GACAGGAGCG ACTTTTCTT

3301 AGCAGACTGA AGAAGATTC CTGATCCCTT GGGTTTCTGC CCTTATTTTG CTCACTCTCT GCGCCAGGCC CATTCGCCCT CTCGAAACAC AGCTGACCA  
TGTCTGACT TCTTCTCAG GATCAGGGA CCGAAGAGCG GGAATGAAC GAGTAGGAGA CCGGTTCGGG GTACCGGAG GAGGTTTGTG TCGACGTCT

3401 AAGGTCACA TTCCAGAAC CCGACGCCA CGAGAGCTGG GAACAGAAA ACCCTGCCA AGACCAAGT CAGTAGGGTC AGCGGAGAC GCGATACAC  
TTCCAGTGT AAGGTCCTG GGTTCGGGT CTCTCGACC CTTTGTCTTT TCGGAGCGGT TCTGTTTCA GTCATGCCAG TCCCGTCTT GCGATACAC  
CCTTATCTG

3501 GCTTAGCTTA GCTGGGAGG TCGAAGAG CAGTGTGTGT CAGGCTCTGA GCGATCCCG TTAATCTCC TTAGGCTTAC TTTTNTAA GTGGACCAT  
CGATCGAT GCAACCTCC ACTTTCTTTC GTACACACA GTGGAGACT CCGTCAGGC AATGAGAGG ACTCGAGTG AAAATATTT CAGCTGCTA

3601 GGTGCTTGC CTATCAGGT GTTAGAGAT TCGGTGAGT AGACAGACA AAAGTTTTC TCGCTGAGT AGCTTCCAC TCATTCCTT AAGCGTAT  
CCAGGAGG CAGTATGCA CACTCTCTA AGGCTCTGA TCTTGTCTT TTTCGAAAG AGGAGCTCA TCGAAGGTG TCGGCTATA TTGCGCATA

3701 GCAATCTAG TTTGATCAG CTAGTCTCT GTCCATCTT ACCCGCGCT TCGAATCTG ATTTTGGG CAGAGGGG GGTTCGGG GAGCTGCA  
CTTAATGAC AAATCTCTT GATCCAGAA CAGGTGAGA TCGCGCGA ACTTAGACC TAAACCCG GTCTTCCG CCAACCGCT CTGACCGTT

3801 GCACTTTGG GAGGTTTTC TTTCTCTC ATAAAGAAC AAAGCTCAT TCTGCGCTC TCGTGTCTT CTCTAGCTG GGTGTACAG CATAGAGT  
CTGAAACCC CCGTCAAG CCGTCAAG TATTTCTT TTTCGAAAT TTTGAAAT TCGTGTCTT CTCTAGCTG GGTGTACAG CATAGAGT  
GTATCTTCA

3901 AGTGGTTCAG AGTCATCTT TCTTCTTA TTTTCTTG ATTTATTTT TTTATTTTT GTGTGAAGT GTCTGCTAC AGTGTCTT GTGACACA  
TCAACATCT TCGATAGA AGAAGAAAT TTTTCTTG ATTTATTTT TTTATTTTT GTGTGAAGT GTCTGCTAC AGTGTCTT GTGACACA  
CACTGTCTT

4001 TCGATCTTT GTGTCTATG AGTCAGAG AGGCTTTGA ATGCTCGA ACTGAGTTT TGAACGTTA TGAAGTCCG TGTGATCTT GAGATCAA  
ACGTACAGAA CACAGTAC TCGATCTT TCGGAACT TATGGACT TGAAGTCAA ACTTGTCTT ACTGAGCTG ACAGTACGA CTCTAGTTT

4101 CCGAGTCTT CTGTAGAAC AAGTCTCTT AAAGCTGAG CCACTTTCC AGTCCAGAG CCACTCTG AGGCTTTC TAACTATG ATCTCGGG  
GGTTCAGGA GACATCTTG TTCTAGAA TTTCGACT GTTAGAGG TCGAGGCTC TCGAAGGAC TCGAAGGAC TCGAAGGAC TCGAAGGAC

4201 GACCACTTG GCAACCTT CAGTACCTT ATTTATTTA AAAAAAAT GCACTCATG GCACTCTT CTAGACTC ATACTAAGT GATTTCTT  
CTGTGGAG CCGTGTGAA GTACTGAG TAAATAAAT TTTTCTTTA CCGTGTGAC CCGTGTGAA GATCTGAGT TATGATCTT CCGTGTGAA

4301 ATAAAGAGT GCTACTTGG GTAGATGCC AGGTTTGGG CCAATTTCA AGCACTGCA CACTTCTGA GCGCTCTT TTTCTTCT GTATCTAG  
TATTTCTT CAGTGAACC CATCTAGG TCGAAGACC GGTTAGGT TGTGACCT GTGAGACTT CCGGAGGCA AAGACAGA CATGTGTG

4401 GCGAGGTCG CTTCGTGTC TCTCTCTAT GCGCGGCT AGTCTAGCG CCAATTTGA ACATTAAT TACTCCCTA CAGAGCGTG AAGCTAAGT  
CGTTCAGG GAAACACAG AGAGAGATA CCGCGCTCA TCGAGTGC CTTTACTT TGTGATTTA ATGAGGAT GTCTGCGAC TTGAGTTA

4501 GGAACCGGC ATTAAGGTC TTAGAATC TCACTGCA TTCTTTAC ATCGGAGGG GAGGTGATA CATGTAGCA GCTGCTTC ACTTTTGG  
CCTTTCGG TATTTCCG AAATCTTAG AGTTAGCTT AAGAAATG TAGGCTTCC CTCACCTAT GTACATCTT CCAAGAGG TGTAAACCC

4601 GAGCGAGCG AGCGTAGGA AATGAGAC AGCTCTTAC AGGCTTCTT ACAGCATCTT GCAACACAC AAGCGGAG TCGGAGAG AGCGGAGC  
CTGCGTCC TCCATCTT TCGCTCTG TCGAAGAT TCGGAAAG TGTGTAGAA CCGTGTGTG TTCCCTCTG AGCGCTTCC TCGGCTTCC

4701 AGGTTGGG GTGCTGGAG AGCTGGGA GCTTGGGC TCGTTCGGG GCGAGCGCG TGAAGCTAG AGCGCGCG TCAATCTT GACTTCTG  
TCAACCGG CAGGAGCTT CCGAAGCGG AGCGAGCGG CCGCTGGGC ACTTGGATC TCGGCGCGG ATTTAGAA CTGAGAGAC

4801 CTCAGAGCG TGTGCTCT TTAGCACTT AGCTCCCG TCGTGAAT GAGCAGCGG TTTGTTCGG CCAAGCGGT CTCTAGCTT CCGCTCTG  
GATCTTCC ACCAGACA ACTGTAGAA TCGAGCGAC AGAATCTA CTTGTTCGG AAACAGCGG CCGTGGCGA GAGTGGAG GCGCAGAC

4901 TCGAGCTTC TCTTCCCTT CAGTCCCTT CTAGTCTT GATTCGGA CCGGCTCC TCTTCTCT TCTACTTG TAGGCGACA CCGTTCGG  
AGTACGAG AGAGAGGA GTAGCGAG GATTCAGGA CTCAGGCT CAGCGGAG AGAAGACA AGATGTAC ATCGGTCTT GGAATCGCC

FIG. 3C-2

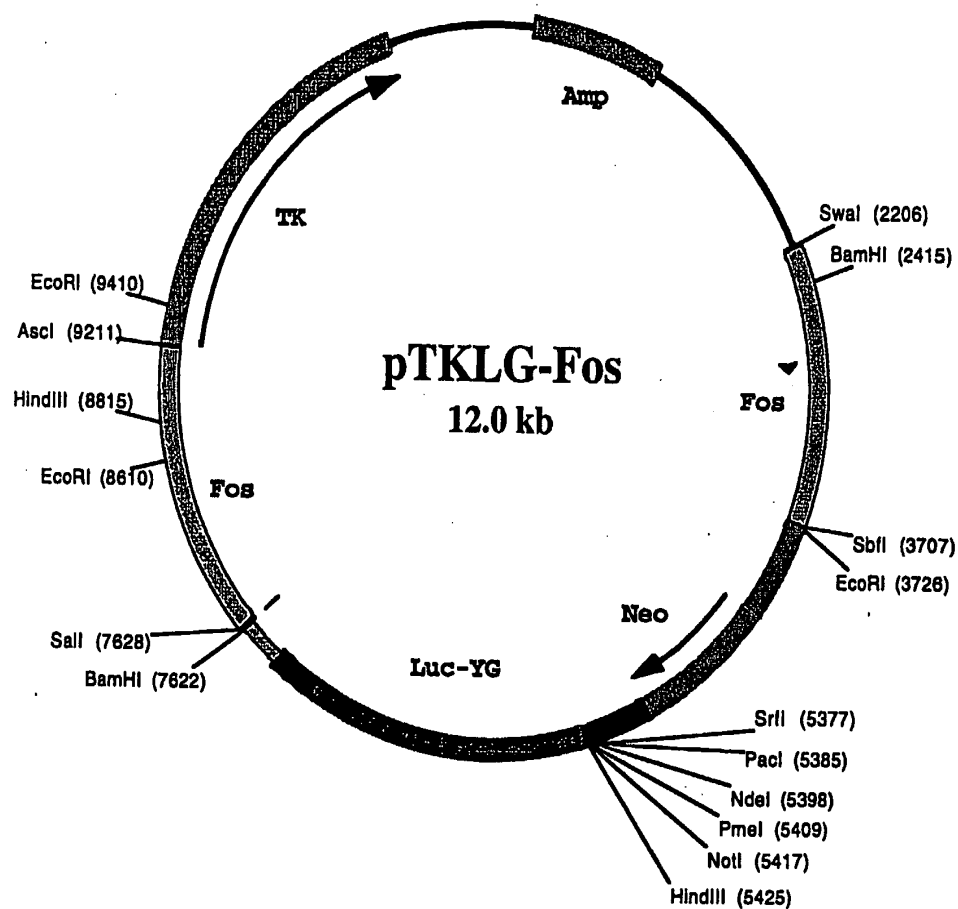


FIG. 4A

1 GCACCTGACC AACCTTCCG GATGCCCGTG CAAAGTATAT ACCCGGTGCT TAGCAGAGCG TGAGAACTTT TAGCCGAAAG CCGGCTCCCT AAGCCGAAAGC  
 COTTCAGCCCG TTTCGACCCG CTACGCCGAC GTTTCATATA TGGGCCACCA ATCTCTCTCG ACTCTTGAAA ATCCGCTTTC GCGCCGAGGA TTCCGCTTTCG

101 TAGCCAAATA GCGGAGAAAG AAGAAAAAAG AAATTCGAGA GAAGCTTCCA GAGCTCTCTC CTCTTCCCTC TTCTTTCAAA AGGACTGCAA GTCCCGAGTGC  
 ATCCGTTCAAT CCGCTTCTTT TTCTTTTCTT TTCTAGGCTT CTTCGAGGAG GAGAGGCGAG AAGGAAATTT TCTCTAGCTT CAGCGGTCTAG

201 ACCCTCCACC CAGCAAGAGT TAGGCGCTCG AACCCCGGTC AGCTCTGCTC GCGCTCTGCG GGAAGTTAAC GCGGAGCCCG TCGGTAAAGC GTGAGCGCTT  
 TGGGAGGTGG GTCTTCTCA ATCCCGGAGC TTGGGGCCAG TCGAGAGGAG GCGGAGGAGC GCTTGCAATG CCGCTTGGCG ACGCATTTTC CACTTCCGGA

301 GGAATCTCC GTCTGACCGC GCGGACCGAC AGCCCGCAGC CCGTCCCGCC GCGCCCGCCC TGACGTCCCG CAGCTTCTTA TTTTGGAGC CCGAGCGCCG  
 CCTTAGAGAG CAGACTCCGC CCGGTCCGTG TCGCGCTTCG GCGAGCGCGG ACTGCAAGCC GTTCGAAATG AAGACTTTC GCTTCAGATG GCTTCCGTG

401 GTTCTAAGAG GAGCGCGCAG COTTCGCTTG TGATTTGCGT TCGCGCGCAG CTTCAGCCAA TGACGCTTCC CTTCCTATTG GTAGAGCGTA GCTCCCTTCC  
 CAGGATTTCC CTCCCGCTTC GCGGGAAGAC ACTAAGCCAG ACCCGCGCTC GAAATCGGTT AGTCCGAAGG GAGGATATA CATCTCCATG CAGAGGAGG

501 TTCTTTTTCG TGTCTCTTCC COTTCGCGCG GTCTCCAGGA GAGAGAGCTAG GATCTCTTTC GCGATCGGGA CTGTTTCTCA CCGCATGCTC TCGGAGACT  
 AAGGAAAAAC ACCAGAAAG CAGCGACCCC CAGAGGTCTT CCTCTGATCG CTAGAGACAG CCGTAGCCCT GAGCAACAGT GCGGTACGAG ACGCTCTGA

601 TGTGTGAGAG TGTCTCTTTC TCATAGAGTA GAGGCTTTTG CCTAGAGTTT AGCGCTCTTA AGCGGAACT GAGGCGCTCA TCTCTCTCAG GCACACATAT  
 ACACACTTGG ACAGGACAC TTCCGAAAGC GAGCTCACAA TCGCGAGAT TCGCGGAGAT TCGCGGAGAT CTTCGAGAT ACGAGAGCT COTGTGTATA

701 ACCTGCTCTT GAGCTCTGTA CACTCAGTCC TTCCGAGGTG TTCAAACT AGATAGCTTA GCTTACCGAG AGGAGCGAG GTGTCTCTA AAGGCTCTGC  
 TGCACAGGGA CTGAGATCT GTAGTCTAG AGGCTCCAC AAGTTTGTGA TCTACTGAT CCGATGCCCT TCGGTCCGTC CACGAGAT TTTCAGAGCG

801 CTTCCTTAG TTCCAGGCT CTGATTTGCC AGGATTCAG CCGTTCCTTC GCGACGCCCC CTAGATAGT TAGGCTCTA GATTCGACT TCGCGAAGG  
 CAGGAGATC AGGTTCGGA CACTAAGCG TCCGTAGCTC GCGAAGCGAG GATCTCATCA ATTCGAGAT ATTCTGAGAT AGGCTCTTC

901 GCGCGCGCGG GCGCGTGTG GAGCTCTTCT GCGGACCGAG ATCTATATTC ACCCGATCCG CTCCAGGACA GTCTAGGAGA TTCTCCGCTT CACTTTTCTC  
 CCGCGCGCGG CCGGACTTTC CTGCGAGAA CCGCTCTGCT TAGGATAGAG TCGGTAGGAG GAGCTTCTCT CAGACTCTCT AAGAGCGACA GTGAAAGAG

1001 TCGCTCTCAG TTACTGAAA COTCTCAGTC TCAGCTGGGA GAGACAGACA CTGCGAAGCG ATCTCTCTCA CTCTTAGGCT GCGCGCGTA CCGCTTCTGA  
 ACGGATGTC AAGTACTTTT GGCAGTCTG AGTGACCTTT CTCTGTCTGT GAGCTTCCG TAGGAGATTT GAGATTCGCG CCGAGCGGTT GTGCGAACT

1101 ACTGGGATCT CCGCTTCCCG GAGCGCTCAT GCGATCGCGG GTGTGTGTGT GTGTAGATGG AGAGGAAAGC TTGCGTAGG CCTCTCCCTC TCGCTCCCTC  
 TGACGCTAGA GCGGAGCCCG CTGCGGATA CCGTACCGCC CACACAACTA CACTCTACC TCTCTTCCG AAGCGATTCC GAGAGCGGAG CCGGAGCGAG

1201 GTGTGTGGGG GTTGGCGGCT TTAGCTGTGA TGTGTGTGTG AATGTCTCTG GCTCCGATCC GCGATTTTGT CAGCAGGTTT CAGCTCCCTC CCGCTCCCTC  
 ACACGACCCC CAGCGCGCCA AAGCGACAT ACACACACAC TTACAGACAC CAGGTAGAGG CCGCTCAACA GTGTCCAGG ACAGGTCCGA GAGAGCGGTG

1301 CCGCGCGCGG ACAGCTAGA GTACCAACC CCGGCTGTGA TTACCAACC CCGTGAAGCG TGCAACTTT CCGCGAGGA GAGAGGAGT GTAGAGGCA  
 CCGCGCGCGG TTGGATCTT CAGTGTGTGT GCGCCACACT AAGTGTGTGG CCGCTTGGG ACTTGGAAA CCGCTCTCTC CATCTCTCTT

1401 GTTGAACAGA ATCTCTCTT AACCACTCG TCACGCTGTA GTGGAAGGTT GGTGTGTGTG CTTTCTTCC TGTAGACAC ACATCCACAC CCGCTCACCC  
 CACTTCTCTT TAGAGAGTA TTGTTAGCC AGTCCACAT CAGCTTCCCA CCGACACAC CCGAAAAGCG ACAGTGTGTG TGTAGTGTG GCGAGTGGG

1501 TGTCTCTCT CAGAGGCTCG GTCTCTCTTA TCTCTCTTGG GGTGTGTGTG CCGGTGTGTG TTGTGTGTG TGTAGGCTT GTCTACGCTC GTCTACGCTC  
 ACAGAGTGA GTTGTCCGCG ACAGAGAACT CCGACACACA CCGGACCGGA AACAAACACA CAGATCCGGA CAGACACATA CAGATCCGCG

1601 GTAGAGTGC GCGGCTCTCG GCGAAATGCC CCGCTCTCTC GTCCCAAGCG TCACCGCAT CAGACAGAG CAGGATCTTC AGTGTCTCTT GCGACCGAC  
 CATCTCTAG CCGCCAGAG CCGTTAGCG GCGGAGGAG CAGGTTGCC AGTGGGTA GTTGTGTG TCTCTAGAG TCACCGACA COTGTGTG

1701 CTCTCTCTCT CCGTCCGCGA GTCCGAGCG CAGCGCTTTC CCGTCCGCTC CCGGCGTTCG AGGTGACAA CTGGGATAC TGTAGGCTC TTGTGTGTG AGTGTGTG  
 GAGTAGAGA GTTACCGGCT CAGGTTCCCG GTCCGTCAG CCGGCGTTCG AGGTGACAA CTGGGATAC TGTAGGCTC TTGTGTGTG AGTGTGTG

1801 CCGTGTGTC CTACAGACT GCGCGGCAAG GCGGAAGTGG TCGGCTTCA ACCAGCAAA CCGCAGTGG AGCTGTCTT CCGGCTCCAG CAGAGCGAG  
 CCGACTCAG CATGTCTGTA GCGCGCGGTT CCGGCTTCA ACCCGGAGT TGTGTGTG GTGTGTCTC TGGACAGTA TGGCTCTCT

1901 GCGTAGAGA CCGCGAGAG AGACGTAGG TAGGAGGCT CAGGAGTTGG GATGAGGAG CCGTGTGTG GATGTGTGT CAGTTTGTG AGTGTGTG  
 CCGACTCTCT GCGGCTCTTC TCTGTCTTC ATACTCGGA GTCTCTAC CCGCTCTCTC GGTGTGTG CCGACAGGA GTCAACAGTA TCAGCGAGG

2001 TCGATGCTT GAGATCTCT AGCAGGCTT AGCGGAGG TGCTGTGTA TGTGTGTG GAGCTGTGAC CCGGCTCTC AGAGGTGGA CCGAGGATTC CCGGAGGATTC  
 ACGTAGGTA CTCTAGGA TGTGTGTA TTGCTCTCT CCGGCTCTC CCGGAGGAGT CTGACATG GGTGTGTG TCTCTCTCT CCGCTCTCT

2101 GAGGCGAGC TGTCTCTCT ATGAGTCTTA GCGTCTCTCT CAGAGATCTA TGTGTGTG AAGTGTGTG TCGGCGGAG TCGGTGTG AAGTAGAGA  
 CTGCGTGTG ACAGATGTA TGTCTCTCT CCGGCTCTCT CAGAGATCTA TGTGTGTG AAGTGTGTG TCGGCGGAG TCGGTGTG AAGTAGAGA

2201 GAGGCGAGC TGTCTCTCT ATGAGTCTTA GCGTCTCTCT CAGAGATCTA TGTGTGTG AAGTGTGTG TCGGCGGAG TCGGTGTG AAGTAGAGA  
 CTGCGTGTG ACAGATGTA TGTCTCTCT CCGGCTCTCT CAGAGATCTA TGTGTGTG AAGTGTGTG TCGGCGGAG TCGGTGTG AAGTAGAGA

2301 TGTGTGTG AAGTAGAGA AAGTGTGTG AAGTGTGTG AAGTGTGTG AAGTGTGTG AAGTGTGTG AAGTGTGTG AAGTGTGTG AAGTGTGTG AAGTGTGTG  
 AAGTAGAGA TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

2401 AAGCGAGAG GTTCCAGAG AGCGAACAA CCGGCTCTCA GCTAGTCTA GCTAGTCTA GCTAGTCTA GCTAGTCTA GCTAGTCTA GCTAGTCTA  
 TTCTCTCTC CAGGCTCTC TCGGCTCTC TCGGCTCTC TCGGCTCTC TCGGCTCTC TCGGCTCTC TCGGCTCTC TCGGCTCTC TCGGCTCTC

2501 CCGGCGGCTT CTGAGCGCG TCGTGGGAG ACTCTGCTT GTTCTTCCCG GGTCTCTCAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG  
 GCGGCGGCTT GAGTCTGCG AGGCGCTCG TCGAGCGGAG CAGAGGCGG GCGAGGCGG GCGAGGCGG GCGAGGCGG GCGAGGCGG GCGAGGCGG

2601 CAGGCGGCTT TGTAGGCTA TGTGTGTG CAGTGTCTA GCTAGGCTT ACTTGTGCT GTTGTGCTT TGTGTGTG TGTGTGTG TGTGTGTG  
 GTCCCGGCTT ATCTGCTT ACTTGTGCT CAGTGTCTA GCTAGGCTT ACTTGTGCT GTTGTGCTT TGTGTGTG TGTGTGTG TGTGTGTG

2701 ATGTGACCT CCGGCTCTT TCTGAGCTA AGGAGACAG CTAGGAGGAG TGTGTGTG ACTTCTCTT CTTCTCTTA TGTGTGTG TGTGTGTG  
 TGTGTGTG GAGCGGAG AGATGCTG TCTCTCTT GATCTCTCT ATTAGAGAG TGTGTGTG GAGTGTG TGTGTGTG TGTGTGTG

2801 CCGGCTCTT TTTTCTCT TGTGTGTG ATCTAGCTT GTGTGTG CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

2901 CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

3001 CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

3101 CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

3201 CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

3301 CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

3401 CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

3501 CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT CCGGCTCTT  
 GAGCGGAG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG TGTGTGTG

FIG. 4B-1

3601 CAGTTGGAG GAGCTGTCC AGCCCCCTGG ATCAGCAGCA AGAATGTATG AGTGTGGGTG TGGGCGGTG AACCTACTCT GTTGTGTCC TGACACCAAA  
GTCAGCCCTC TCGGAGACG TCGGAGACG TAGTGTGTGT TCTTACATAC TCAACGCGCA ACCGCGCCAC TTGGATGAGA CACACAGCGG ACTGTGTGTG

3701 TTTCTCTTTC TCTGTCTCT ATGACCTGCG CTTGCTGGGA TCCATATAGA AACTGATCAG CTTGAGAGAG AAAGGCGAGA CTTGAGTTCG GAGATCCCGG  
AAGAGGAAG AGACAGAGGA TACTGGACCG GAGCGACCGT AGGTAACTCT TTGACTAGTC GAACTTCTCC TTTTGGGTCT CTAAGCTCAG CTTTACCGCG

3801 AGCTGCAAAA AGAGAGGAAA GCGCTGGAGT TGTGTCTGTG GCGCCACAAA CCGGCTGCA AGATCCGCTA CGAAGAGGCG CCGGCGCGAG GCGGCTGTGG  
TGAGGTTTT TCTCTTCTCT GCGGACCTCA AACAGGACCA CCGGCTGTGT TCTAGGGAT GCTTCTCCCG GCGGCGGTC GCGGCGAGCG

3901 CGAGGTGAGA GATTTCGAG GGTTCACATC CGTTCAGGAA GAGGCTTTCG CTTGCTGTCT GCGGCGGCTT CAGACGCGCG CCGTGGCTCT CAGAGCGAGC  
GCTCAGCTCT CTAAACGGTC CAGTTGTGAG CGCATTCCTT CTGCGGAAGC GAGCGGACGA CCGGCGGCGA GGTGTGTGCG GCGGCGGAAA GGTCTGTGTG

4001 CGAGAGCGAC GCGGCGGCTT GAGGCTTCTT CTCTTTACAC ACGTGTAGT TCAAGTCTCT GCGGAGCGCT TCGGCTGTGT TAGGCTTTCG TACACTTCTT  
GCTCTGTGTG GCGGCGGCTT CTGCGGAGGA GAGGAATGTG TGTACTTCA AGTTCAGGAG CCGCTGTGGA AGGCGGACCA ATCGGAGAGC ATGTGAGAGA

4101 GATTGTCTCT CAGCTGCGCG GAGGTCTCGG CGTTCGCGCG GCGGCGAGCG ACCAGCGGCA GCGAGCAGCG GTCCGAGCGG CTGAAGCTCG CCGTCTCTCT  
GCAACAGCA GTGAGCGCG CTCCAGAGCG GCAAGCGCGG CCGGCTGTGT TGTGTGCGGT GCTCTGTGTG CAGGCTGTGG GACTTGTAGG GAGAGGAGGA

4201 TGTCTGTGAA ACTCTTGTAGA CAAACAAAC AAACAAAC CAAAGGAGGA AGATGAGAG AGATGAGAG TCTATCTCTC TGTCTGTGTA TGTCTGTGTA  
ACAGAGACTT TGAGAAATCT GTTGTGTGTG TTTGTGTGTG CGTTCCTTGT TCTCTCTCT TCTATCTCTC TGTCTGTGTA TGTCTGTGTA TGTCTGTGTA

4301 GTTGTGAGCG TTTGACTCTT CTCTCTGACC AGCTGCGCGG TCTGCGCTAG GAGATGAGCG AGAGAGCTCC TTTGTGTGTA TGTCTGTGTA TGTCTGTGTA  
CAGACCTGGG AAATGAGAA GACAGACTGG TGAGCGCGCG AGAGCGTACG CTGTACTGCC TTCTGTGAGG AAACAAACAA CAGAGAGAGT TGTGTGTGTA

4401 TGTGCGCGCG CAGAGCGGGA GAGCTGTGTA CTTTGGGAG AGGCGGTGCG GCGGCGGAGT ACGCGCTCTT TGTCTGTGTA TGTCTGTGTA TGTCTGTGTA  
ACAGCGCGCG GCTCTGTGCT CTGAGCGACT GAAACCGCTG TCGGCGGCGG CCGGCGGCTT TGTGTGAGAG AGTGTGAGGA ACAGAGCAAT GAGTGTGTA

4501 ACTTCTGGCG ATGAGTGTCT GACTGTGTGG GTGCGGTGCG GTCCAGCGCG CAGCTTGTGG GCTCTACTGT AGGCTGTGAG GAGAGAGAGT CTGAGTGTGG  
TATCTGTGGA CTGAGCGCG CAGTGTGCG CAGTGTGCG CAGTGTGCG CAGTGTGCG CAGTGTGCG CAGTGTGCG CAGTGTGCG CAGTGTGCG

4601 GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT TGTGAGGT TGTGAGGT TGTGAGGT TGTGAGGT TGTGAGGT TGTGAGGT TGTGAGGT TGTGAGGT  
CCAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

4701 CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA  
GTTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

4801 GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA  
CTCTGTGAGT CCGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA GAGAGTGTGGA

4901 TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA TCAAGTGTGGA  
AGTGTGAGAG GAGAGTGTGGA TTTGAGCGCG TTTGAGCGCG TTTGAGCGCG TTTGAGCGCG TTTGAGCGCG TTTGAGCGCG TTTGAGCGCG TTTGAGCGCG TTTGAGCGCG

5001 TATTATGCTA TATTATGCTA TATTATGCTA TATTATGCTA TATTATGCTA TATTATGCTA TATTATGCTA TATTATGCTA TATTATGCTA TATTATGCTA  
ATGATGAGAT ACGAGGAGAG CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5101 CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA  
GTTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5201 CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA  
TGTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5301 TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT  
AAAGAGAGAA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5401 GAGTGTGGA GAGTGTGGA GAGTGTGGA GAGTGTGGA GAGTGTGGA GAGTGTGGA GAGTGTGGA GAGTGTGGA GAGTGTGGA GAGTGTGGA  
CTTCAAAAGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5501 ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA  
TAGGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA ATCCAGTGTGGA

5601 AATATGTGGA AATATGTGGA AATATGTGGA AATATGTGGA AATATGTGGA AATATGTGGA AATATGTGGA AATATGTGGA AATATGTGGA AATATGTGGA  
TGTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5701 CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA  
GTTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5801 GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA GTGAGTGTGGA  
GTTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

5901 TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT  
AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA AACAGTGTGGA

6001 CAGTGTGGA CAGTGTGGA CAGTGTGGA CAGTGTGGA CAGTGTGGA CAGTGTGGA CAGTGTGGA CAGTGTGGA CAGTGTGGA CAGTGTGGA  
GTTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

6101 GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT GGTGTGAGGT  
CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA

6201 TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA TAGGCTGTGGA  
ATCGGAGGT GGTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

6301 CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA CCGGCTGTGGA  
GTTGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

6401 GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT GCTGTGAGGT  
CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA CCAGTGTGGA

6501 AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT AGTGTGAGGT  
TCACTGTGGA TCACTGTGGA TCACTGTGGA TCACTGTGGA TCACTGTGGA TCACTGTGGA TCACTGTGGA TCACTGTGGA TCACTGTGGA TCACTGTGGA

6601 TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA TGGAGTGTGGA  
ACCTCTTAA AGTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT TTTGTGAGGT

6701 GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA  
CATGTGAGGT GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA GTAAATGTGGA

6801 GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA GAGGCTGTGGA  
TACAGGAGGT TACAGGAGGT TACAGGAGGT TACAGGAGGT TACAGGAGGT TACAGGAGGT TACAGGAGGT TACAGGAGGT TACAGGAGGT TACAGGAGGT

6901 ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT ATGAGGAGGT  
TACGCTTAA GTGAGGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

7001 CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT CTAGAGGAGGT  
GACTGTCTCT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

7101 TBAATGTGGA TBAATGTGGA TBAATGTGGA TBAATGTGGA TBAATGTGGA TBAATGTGGA TBAATGTGGA TBAATGTGGA TBAATGTGGA TBAATGTGGA  
AATGTGAGGT CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA CCGAGTGTGGA

FIG. 4B-2



7201 CTGGTTCGCC CCAATCTGCC TCTTCTCTCA GCTTGTCTTC CATTAAGTGT GTGTCTCAAG GCGACCTGCG TCAGGACTCG TTGTGAGAGC AGCTCTGTGT  
GAGCCAGAGG GTTAAGAGAGG AGAGAGAGGT GGAACAGAGG GTAAATTCACA CACAGAGTTC CGGTGGAGAG AGTCTGTAGG AACACTCTGC TTGAAGATTC

7301 CTGAGTTTCA TTAATAACAC AATTCGCTCG TCGGCTGTCT TCTCCACTCG CTGAGTACC TCAAAAGACC AGCGCTBAAG GTGTGATGAC AACTCTATCC  
GAGCTCAAGT AATTTTGTGT TBAAGGAGC AGCGCAGAGG AGAGGTGAGC AGTGTATCG AGTTTTCTGG TOCGGATTCG CACACTAGTG TTAGAGTAGG

7401 CCAATCTCTC TCCAGCGCAG AGACAGGACT GAGCGGAGGT GAACAAATGA ACAAAMATGA CTAAATATGC ATCGGTGATT AAATCATATA AAGAGCAGAT  
GATTAATGAG AGGTTGCGTC TCTGTCTTGA CTGCGCTTCA CTGTCTTACT TGTTTTACT GATTAATAGG TACGACTATA TTATATATAT TCTGTCTTGA

7501 GACTGATAGA CCAATCTGTT TACGAGAGCA CACCAAGATC CTAGAAATTT GTAGACTAAT TAAATTCAT GTTTCAGATG CATTTGCTCG GAATGTCTGT  
CTGAGCTACT GTTTATGCA ATTCCTCTCT GTGTTCTTAG GATCTAAAA CCTCTGATTA AATTTAGGTA GTTATGAGT GAAACTCTAC GTTAACCGAG CTTTAGAGAC

7601 GGAAGAAAA AAGTGTAAAT ATGAGAGAGG AATAAATGAG AATAGCGGTG GCTTCAGAGA GTTAACTGCG GCGCTGTGCG CTTTGTATCA AGAATGTGAA  
GCTCTTTTTC TCACTATTA TACTTCTCTC TATTTACTC TATCCGCCAC GTTAGCTCTT CCAATTTAGG GCGACCGAG CTTTGTATCA AGAATGTGAA  
TCTTCACTCT

7701 TTGAGCGGAG CAAATGTGGA TACGACTCC GCGCGGAGAG GTGGAATTGA ACCACTCTGT GCTTAAACAG CTACAGGTTT GAAGCTTCCA GCGCAGGAG  
AAGTCTCTTC GTTTATGCGT ATCTATGAGG GCGCGCTTTC CACTTAACT TGTGTAGACA GCGATTTGTC GATGTCCAAA CTTGTAGAGT GCGGTCTGCT

7801 CTGAGATCA TCGCGCGGAA AGGAGCTATT TTAAGTATGT TATAAAGAG GAGATACTA CTACTTTTGA CACTTATGTT CATATTGT GTATATGAT  
GACTCTAGT AGCGCGCTT TCGTATGA AATCTAATCT ATATATTTTC GCTCTATGAT GATGAAAAAT GTGATATCCA GTATATGA CCAATGTGCA

7901 AGATATATTA TTCAATGCT TCGAACAT TTTTTCCTACT TTTTCTGTGT AACATGTGTT TCGTCAATGA AGTGTTCGTT GAATGACTCT ACTTACTTAA  
TCTATTAATT AAGTATGCA AAGCTGTGAA AAAAAGATGA AAAAAAGAC TGTATACAAA AGGATCTATT TCAAGAGGCA CTACTATGAA TATATGATTT

8001 GACTGAATGA CTTCATTTGCT ATAGCGCTT GCAATTTGCG AAGCAGCGCT TAAAGTGTCT TTTTCAAGCA GTCTCCCTTA CTAAAGAGC AATTTTTCG  
TTCAATTCATC GAGTAAAGG TATCGCGGAA GTTAAACCC TTTGTCCCGG ATTTCAAGCA CAGAGGATTT GATTTTCTGC TTAAGAGAGC TTTCACTTTT

8101 GTCAATTTTA TTTTGTGTT TTTTGTGCT TTTTGTGTT TAAATGAAA ACTTCTCAAG GCGCGCTTTC GTAGCGAAT TCGAGATTTT CTGAGCGAA  
CATCTAAAT AACAACAAAC AACAACAAAC AACAACAAAC ATATCTTTT TGAAGATGCG GCGCGGTAGG CATGTCTTA AGCTCAAAA GAGGTCTGCT

8201 GAAGCAGAGC TTCTGTAGG TCTGAGCGCA GCGCGCGGCA GAGCGACAGC TCGGCTGTCT TTTATAGAACT GCAAGTATGT AGGATATCA CTGAGTCTGT  
CTTGTCTCT TAAACATGCG AGACTGCGT GCGCGCGGCT CTGCGCTGTG AGCGCAACAA AATACTTGA GGTTCATCA TCGCTATGAT GACTCAGGAA

8301 AGGTGATGGA GTTGACAGC AACTCGCTT GAGTTTAGAG GCTTAAAGC ATCCCTTTT ATATATATGT GATTAAGGCA GCGGATGAA GCGTCAAGCA  
TGCATGCTT CTGAGGAGG CTCAATCTT GAGTATTTG GAGTAAAGCA CTAAAGGCTT GATTAAGGCA GCGGATGAA GCGTCAAGCA GCGTCAAGCA

8401 TCGATATATC CACAGCGGAG TCTGTGAGC GCACTCTGCT AGCGGAAATG AAGCTTACAG TTGTGTGTTT AATATGCTTG GCGCGCGCT AGTGTGCTTA  
ACCTATATG GTGTGCGCTC AAGAACATG GTTTAGAGG TCGGCTTTC TTTGTATGTC AACACCAAAA TTTATGAGAC GCGGTGCGCG TCAAGCGGAT

8501 TTGAGCAGG TCGGCTTAT TCGGCTGAG TACCTGTGTA GAGAGTGTG TCACTGTGAG GCGAGGTTT GAGGTATGTA TCGTCAAGT TCGGCTTAT  
AAGGTGCTC ACCGAGATTA TCGGCTGAG ATGGAACAT CTCTTCAGAC AGTAAAGCTC GCGTCCAAA CTCTCTGAT TCGTCAAGT TCGGCTTAT

8601 GATCTGCTGCT GTCTGAGAA CGTGTCTTC TCTGCTGCT CTCTGAGTCA AGGTATGAA CTCTCAGCT CTCTCTGAG ACCATATCT CTCTCTGAT  
CTAGGACGGA CAGAGCTCTT GCGAGAGAG GAGGCTTAT TCACTATCTT TCACTATCTT GAGAGTGTG TCGTCTGAG TCGTCTGAG GAGGCTTAT

8701 GCTTCTCTTC TTTCTATGAC GATATGAGC TGTGCTCTG AACTGTAGG TCAAGCGGCTT AATATATGT TTTCTTAT TTTCTTAT TTTCTTAT  
CGAAGAGAG AAGGTATCT CTATATCTG ACGCGAGAG TTTGACATTC AATGCGGCTT TCAATATGTA AAGGAAAT TTTCTTAT TTTCTTAT

8801 TATGATATTA TATGATATTA TATGATATTA TATGATATTA TATGATATTA TATGATATTA TATGATATTA TATGATATTA TATGATATTA TATGATATTA  
ATATATATTA ATATATATTA ATATATATTA ATATATATTA ATATATATTA ATATATATTA ATATATATTA ATATATATTA ATATATATTA ATATATATTA

8901 AAGATTTGCT GACTTTGAA GCAATTTGCT TCGGCTGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG  
CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT CTGAAAGCTT

9001 AAGAGCAGA AATCAGAGT TCGTCTGCT AATTCAGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG TCGGCTGAG  
TCTCTGCTT TTTATGCTCA AAGAAATGGA TCAAGTGTCT AAGGATGTTA GATTTGAGCA AGGTGAGT TTTGTGAGT TTTGTGAGT TTTGTGAGT

9101 TTTGAGGAT GCTGAGGCT CTGACAGCA CACTCTCTCT TGTATATCT CTCTGAGCT GTTGTCTTC GCGCGGATGT CCAATGTGCG GAAAGCTCT  
AACTCTCTTA GAGGTGCGA GAGTGTCTGT GTAGAGAGG ACCAATGAG AAGTCTGAG GAGTCTGAG CAGCGAGG GCGCGGATGT CCAATGTGCG

9201 CATCTGTTC TCAATAGCA CTAGCTATGA AGCTTCCCG AGCTGAGCG GTTTAAATAT TAAAGAGG TCACTTTCTC CCGTCCAGC ACAAAGCAG  
GTAGGAGAG AGTTATGCT GATCGATCT TCGGAGGCG TCGAGTGGG CAAATATTA ATCTTTTCC AGTGAAGAG GAGGCTGCT TGTGTGTTG

9301 CACCAATCT TTTCTATGA CTAGCTATGA ATGAGGCTA ATGAGGCTT TATCTATGAG TATCTATGAG TATCTATGAG TATCTATGAG TATCTATGAG  
GTGTATGAG AAGGTGAT TATCTATGA GATGATCTA TACTTCCAT TATCTATGAG TATCTATGAG TATCTATGAG TATCTATGAG TATCTATGAG

9401 GATCAGAT TCGGAGAGC TGTGTCTCT CTCTATCTA GTCTGAGAG CTTTATGAG ATTTCTCTAT GTTGTGTTA GCGCGGATGT ACAAAGCAG  
CTGTCTCTA AGGATCTCG ACGAGAGG GATGAGGAT CAGAGCGCTG GAAATATG TAAAGAGT CAAACAGCT GCGGTGCTG TGTATTTA

9501 TATTTCCAT GATCTCTAT AACTGTAT TTTCTATG TATGATAT TATGATAT TATGATAT TATGATAT TATGATAT TATGATAT  
ATTAAGGTA CTATGAGT TTTGATTA AAGAGATAC AATCTATCT ATATCTAT TAAACAAA GGTCTATG AATCTATG GAGGCTCT

9601 GTCAATGAG CCAAGAGG TCGGAGAG AAGTATGAA TTTCTGCT ATAGGATCA CAGGAGCAT GATATGAG TTTGAGAG TTTGAGAG  
CATTAAGCT GGTTTTCCG AGGTTGCT TTTCTCTT AAGGAGGTA TCTCTGAT GTCTCTGTA CTTATGTT GCTATGTT GCTATGTT

9701 CTTCTGAG GCGCTGAG CTATGAGC CTACTCAT TTTGAAAT TTTGTGTT GTTGTGTT GTTGTGTT GTTGTGTT GTTGTGTT  
GAGAGAGCT CCGGAGCTC GATGATGTC GATGATGTA AAGCTTTA AACACACCA CACACACCA CACACACCA CACACACCA CCGGAGCTG

9801 GGTGCTGAG ATAGGCGAGT GCTTTAGT TCTCTGAGC GATCTATC CAGACTCTC GCGAGTGC AGTCTATG ATTTCTAT GTATGACTA TGTCTCTA  
CCGAGCTC TATCGGCTA CCAATATC AAGGAGCTG GATCTATC CAGACTCTC GCGAGTGC AGTCTATG ATTTCTAT GTATGACTA TGTCTCTA

9901 GTGCGGCTG CAGGAGAGC AAGGTGAG TAAATTTA AAGTATGAG CAGCTGAG TACAAATAT GCGATTTGA AGTGTGTT GATGTCTA  
CAGGCGAG GTTATGCTG TTTCTATG ATTAATAT TTTCTATG GTGAGCTCT ATTTATTA GTTCAAACT TCAAGGCTA CTAGAGTT

10001 TAACTTAA ATATATGAT TATATGCTG CTCTCTCG GGAAGCAGA AGTTTGAAA TATAGAGT CATGTGCT TCGGCTGCG GTTCTTAC CTAAGGCT

10101 GTAGGAGAG AGGTGCGG GAGGTATG GCGGCTTTC GATAGCAT TGAATGTA ATGAGAAA TATCTATTA AATTTGAA AAAATGTA  
CATCGCTG TCGGAGGCT CTCTATAT CCGTGAAG CCGTATGAG CTTTACTT TACTCTTT ATATATAT TTTTACTT TTTTACTT

10201 CCGGAGTTG GCGTATGCT CACTACTCA ACCAGAGT CATGTGACT TCGTATGAT TCGGCTCTC TCGGCTCTG GTGAGAGA CAAATTTG  
GCGGCTAAC CCGGAGTGA GTATGAGT TGTCTGAG CTACTATG AGGACTCTC GCGAGTGC AGTCTATG ATTTCTAT GTATGACTA TGTCTCTA

10301 AAGTATGCT TCTCTTGA TCTGTGAT TCGAGGAT AGTCTCTA GAACTCGGT CATCAGCT GCGTGAAT GACTACTTA GTTCTCTC AGAGCTCT  
TTCAATCAG AAGAGAGT AAGAGCTA AGTCTCTA CTTGAGGCA GTTATGAG CCGGCTCTA CTGAGTAT GTTCTCTC TCTGAGAG

10401 GTTGTATGA GTATGATCT GCACTTATG CCGGACTTTC GCACTATGA GATAGAGCA TGTCTATAC ATCTCTCA ATATATAT TATCAGAG  
CCAACTAT CAACTATGA GTTATGAG CCGGACTTTC GCACTATGA GATAGAGCA TGTCTATAC ATCTCTCA ATATATAT TATCAGAG

10501 CAGGTATGA GATGCTCAG ACCGATCTC GTTCTGAG CAGAGCTCT CAAATGAG CAACTATGA CAGGCTCTA CAGGCTCTA CAGGCTCTA  
GTCTACTCT CTCTGAGG ATCTATGAG AGGAGTCT GTTCTGAG CAGGCTCTA CAAATGAG CAACTATGA CAGGCTCTA CAGGCTCTA

10601 TCGATGCT GAGGCTCT AAGTATGAT TATATATAT AATATATTA ATCTTAAAA AAAAAAGC AGCGGCTCT GCGGCTCTA CCGGCTCTA  
ACTCTATGA CTCTGAGG TGTATATGA ATATATAT TATATATAT TATATATAT TATATATAT TATATATAT TATATATAT TATATATAT

10701 CAGGCTCT GAGGCTCT AAGTATGAT TATATATAT AATATATTA ATCTTAAAA AAAAAAGC AGCGGCTCT GCGGCTCTA CCGGCTCTA  
GCGGCTCT GCGGCTCT AAGTATGAT TATATATAT AATATATTA ATCTTAAAA AAAAAAGC AGCGGCTCT GCGGCTCTA CCGGCTCTA

FIG. 4B-3

10801 GTCTGGAAA AAAAAGAGA GAGAGGAGG TGAGAGGCA ATAACTDA CATTCCTG OTTGTCTTG CAGTACTA TCTGTAA CAGTCTGC  
CAGAGCTTT TTTTCTCT CTTCTCTTC ACTCTGGGT TATTAGATT GTAAAGAC CACAGAAC GACATCAGT AAGACTATC GTTACGACC  
10901 TTCTGCCAA GGTAGGAGT AACATTTCTT TATTAAGGT ATTGTCTG CTTATTTT CTTTCTAT TATGCTCTG AGATGAG CAGGAGCCT  
AAGAGGTT CATTCTTCA TTATTAAGA ATATTTTCA TAAAGAGC GAAATAAA GACAAATA ATCCAGAC TCTACTTG GTTCTGGA  
11001 TGCCAGCAA GCTAGCTGT TACCACTGA GCGACTCC AGCTTGAC TGGGATTC TAGGAGGG TTCTACACT GAGGACACT CCGACCCC  
ACGTTGTT CCAATGACA ATGTGACT CGGTATGAG TGGAGCTG ACCCTTAG ATCGTTCC AAGATGGTA CTGGTGTGA GGGTGGGG  
11101 ATCTCTCTT GAGATCTT AGGATCTC ATACTAGC TTGATCTT TAGAGGTC TACTAGAC TCAGT  
TAGGAGAGA CTTCTAGA TCGTCAGG TATGATGG AACTAGAA ATTCTCCG ATGATCTG AGTCA

FIG. 4B-4

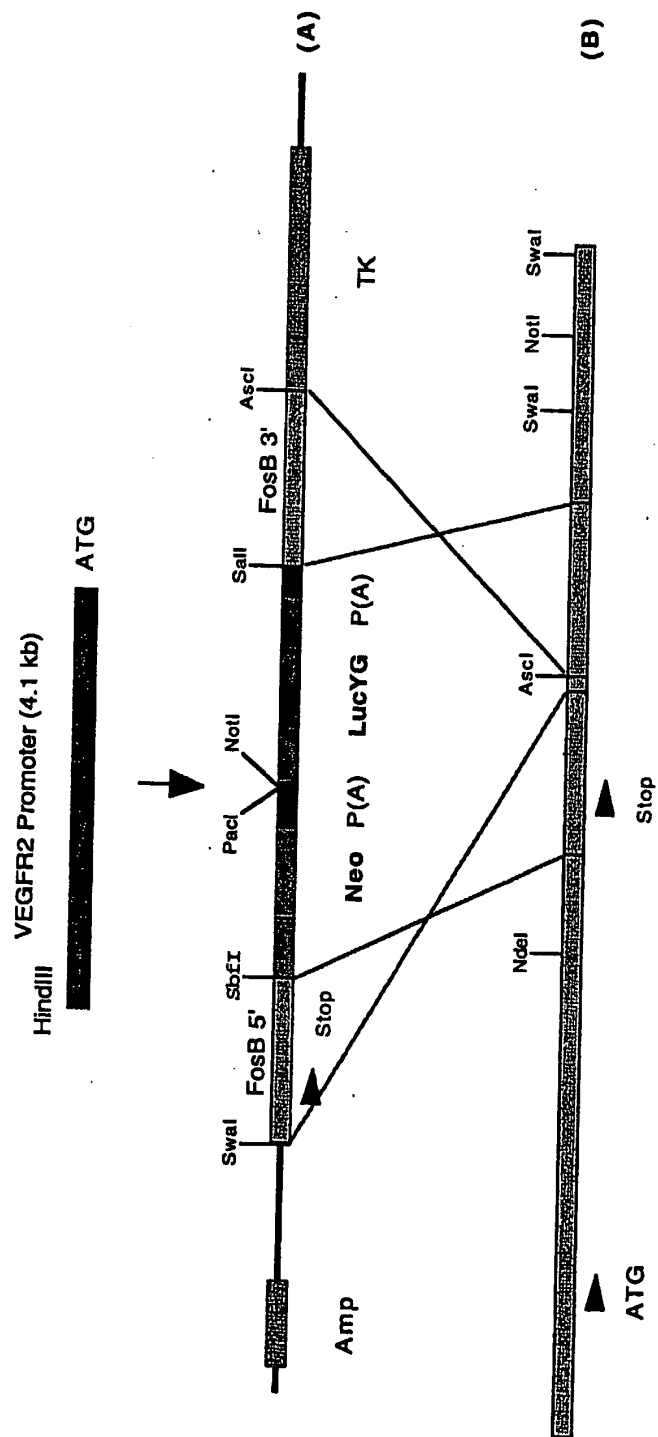


FIG. 5A

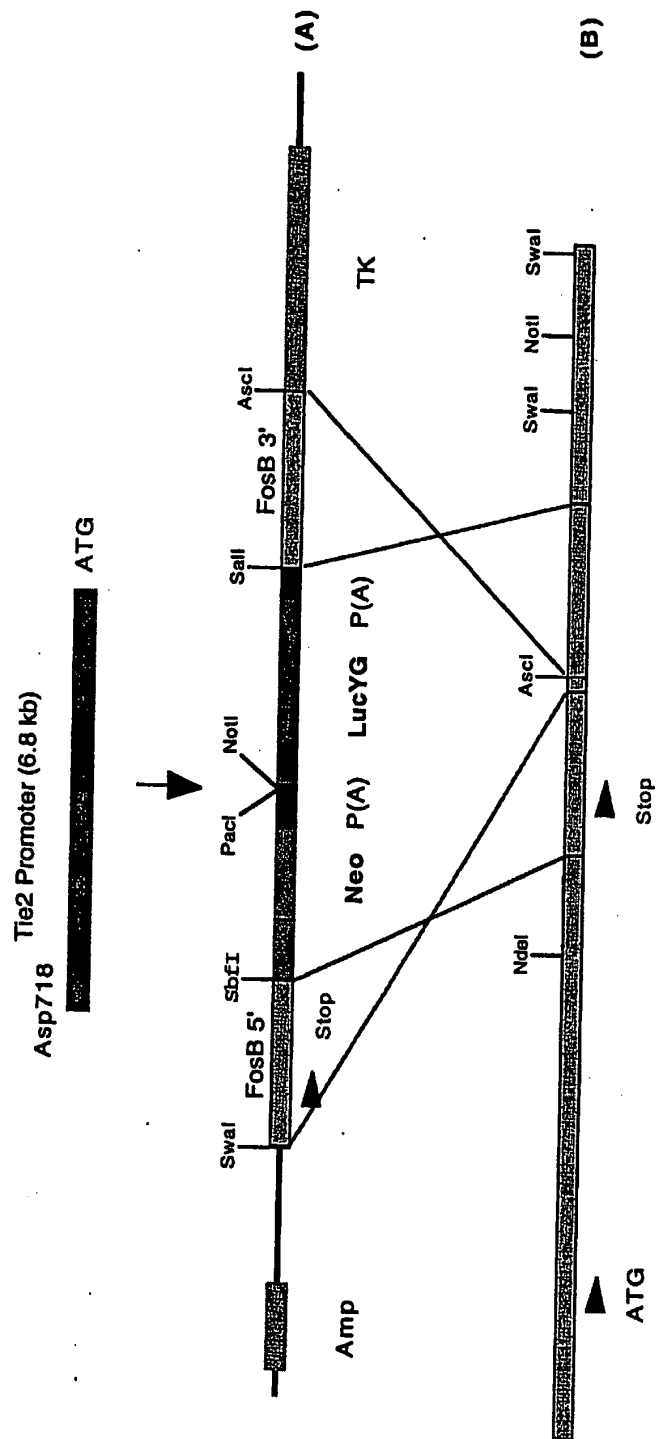
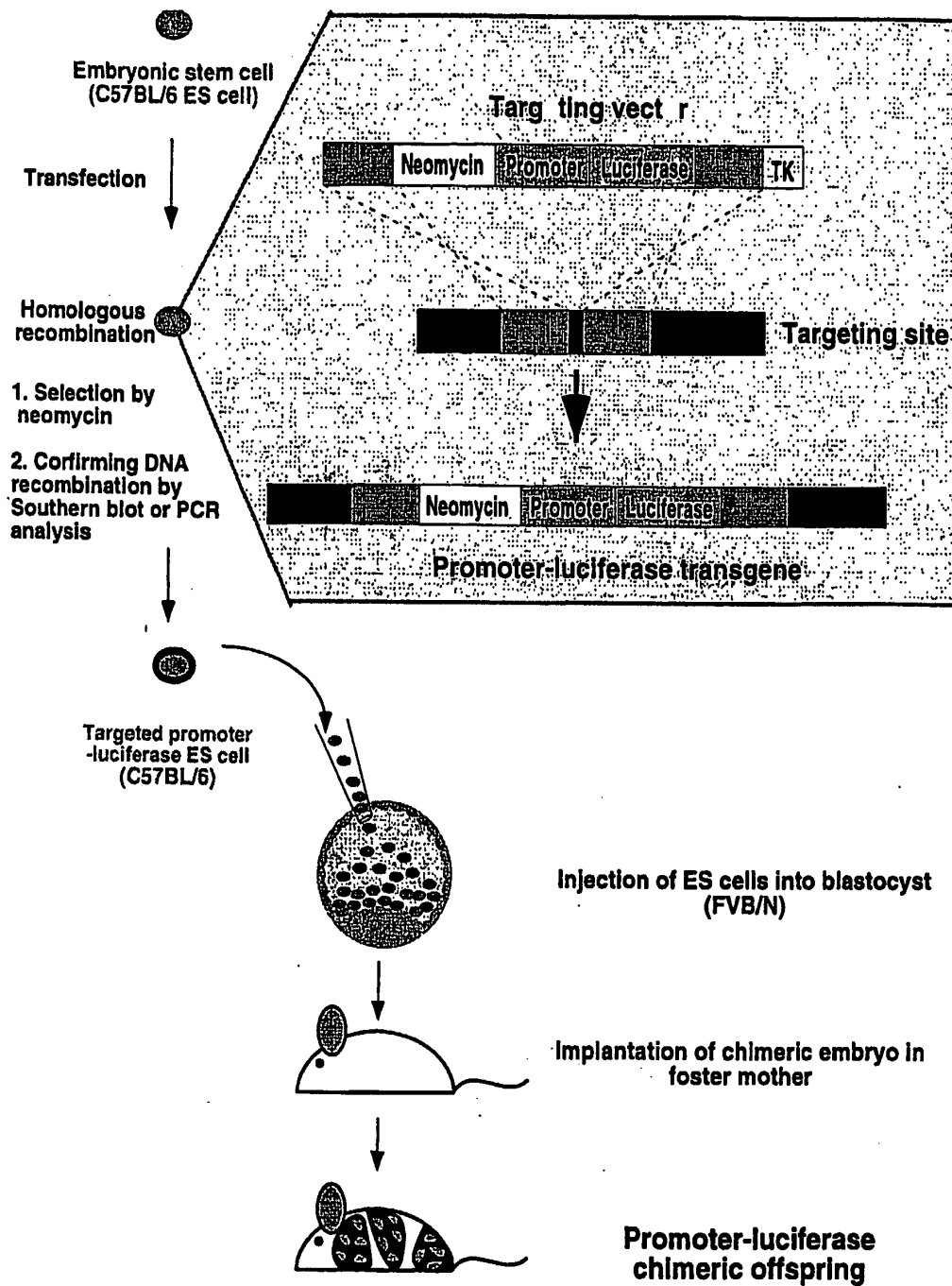
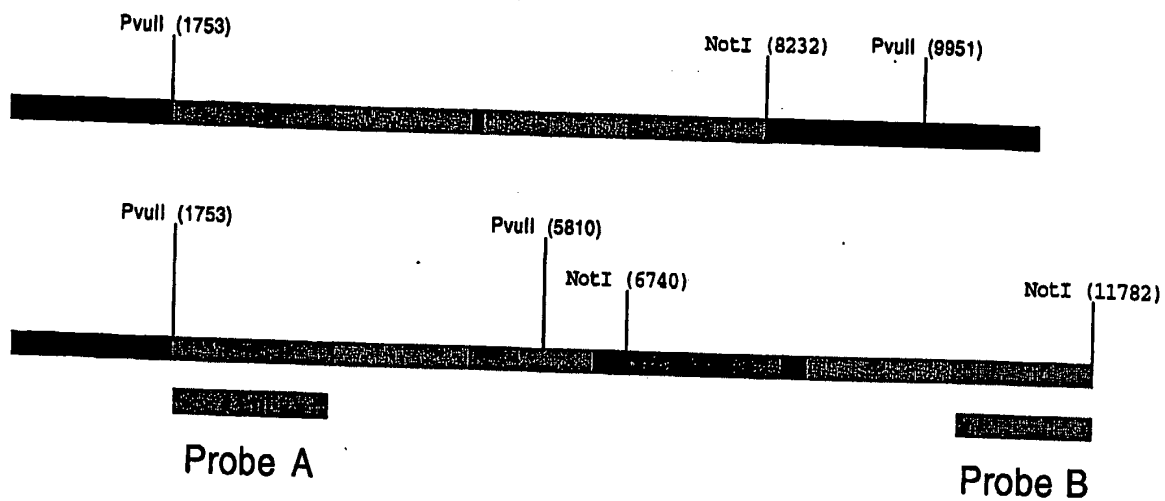


FIG. 5B

<b>VEGF project</b>	<b>VEGFR2 project</b>	<b>Tie2 project</b>
<b>Screening primers</b>	<b>Screening primers</b>	<b>Screening primers</b>
Primers: VF1-VR1A Product size: 0.69Kb	Primers: KF1-KR1 Product size: 0.45Kb	Primers: TF3-TR1 Product size: 0.45Kb
<b>PCR program</b>	<b>PCR program</b>	<b>PCR program</b>
Hot start	Hot start	Hot start
94°C 40 sec 65°C 1 min 30 sec 72°C 1 min 30 sec	94°C 40 sec 58°C 1 min 30 sec 72°C 1 min 30 sec	94°C 40 sec 58°C 1 min 30 sec 72°C 1 min 30 sec
40 cycles	40 cycles	40 cycles
<b>Confirmation primers</b>	<b>Confirmation primers</b>	<b>Confirmation primers</b>
Primers: VF2-VR2 Product size: 0.98Kb	Primers: KF2-KR2 Product size: 0.58Kb	Primers: TF2-TR1 Product size: 0.47Kb
<b>PCR program</b>	<b>PCR program</b>	<b>PCR program</b>
Hot start	Hot start	Hot start
94°C 40 sec 65°C 1 min 30 sec 72°C 1 min 30 sec	94°C 40 sec 65°C 1 min 30 sec 72°C 1 min 30 sec	94°C 40 sec 58°C 1 min 30 sec 72°C 1 min 30 sec
40 cycles	40 cycles	40 cycles

FIG. 6

**FIG.7**

**FIG.8**

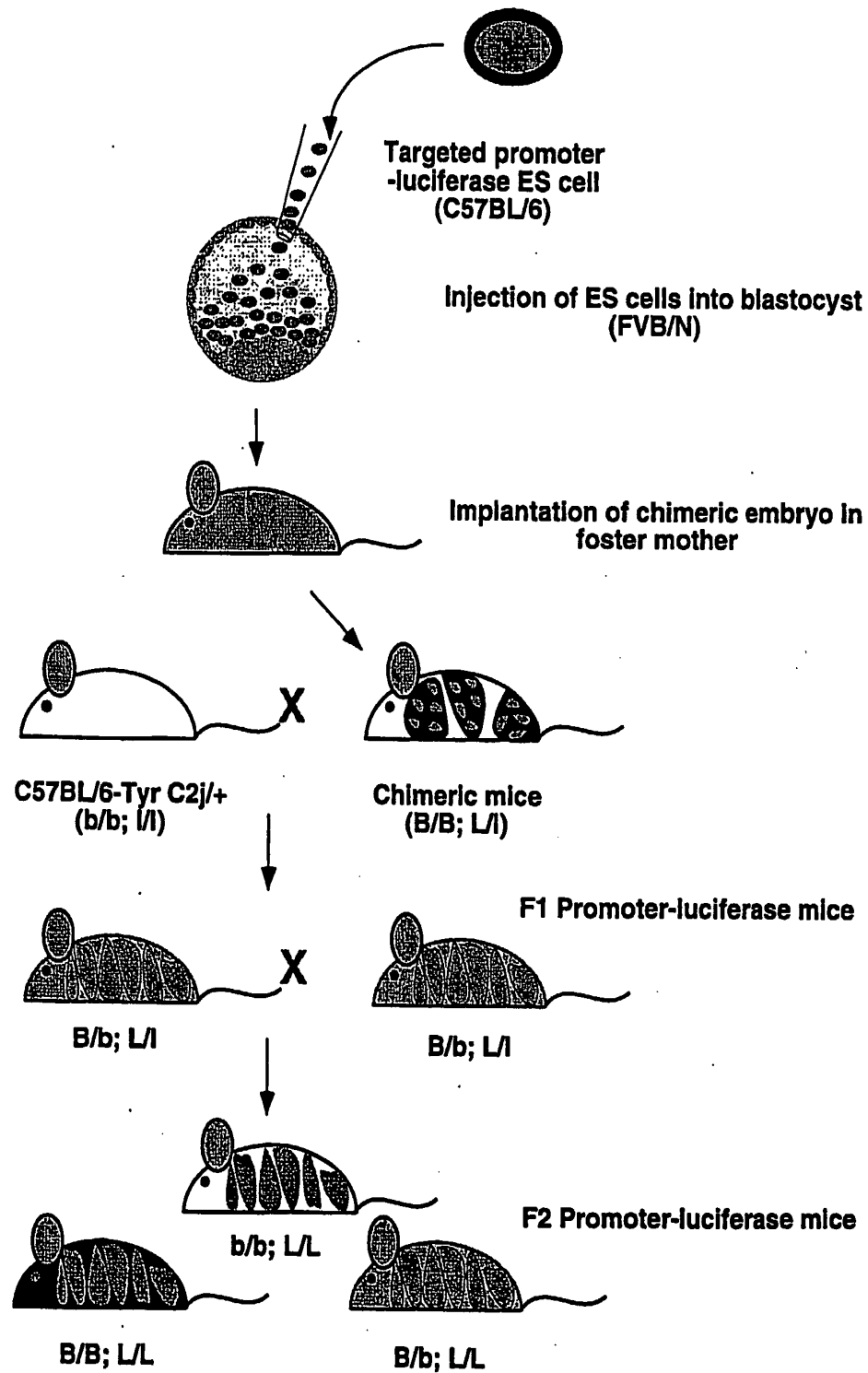
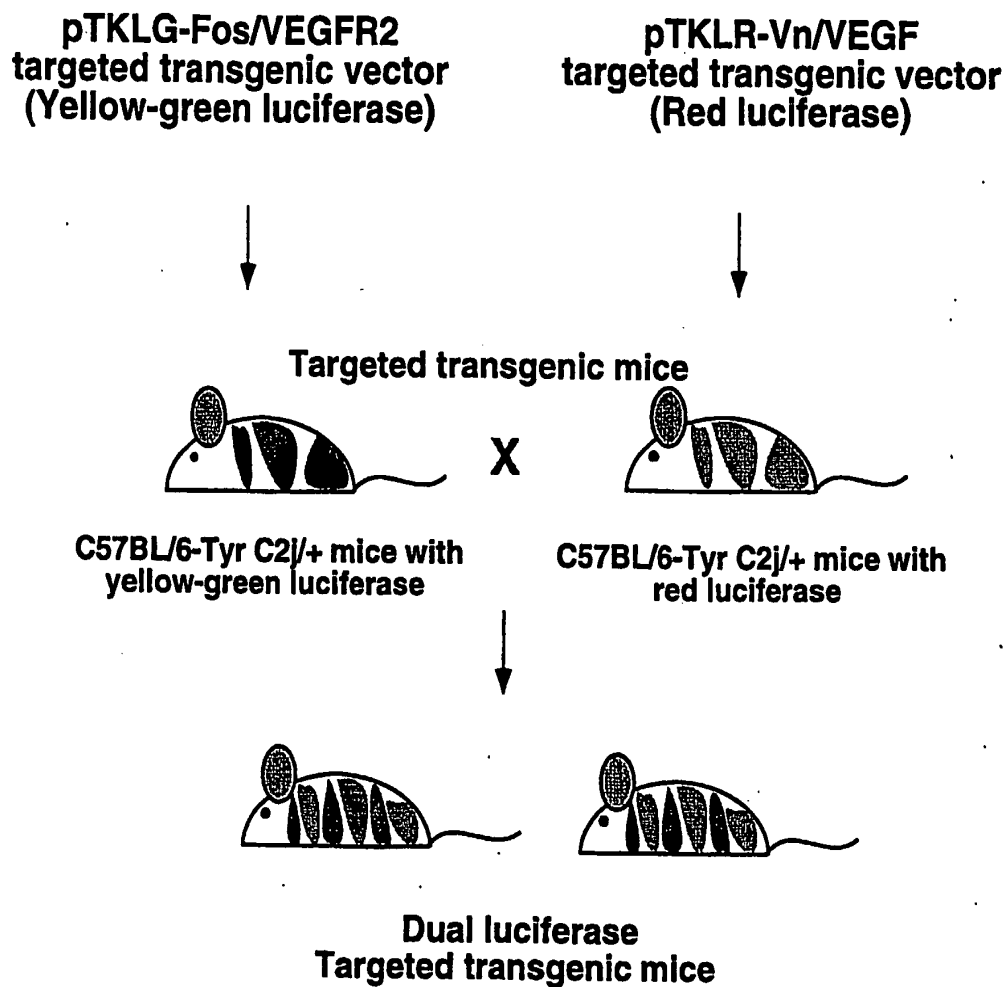


FIG.9



*FIG. 10*

## SEQUENCE LISTING

<110> XENOVEN CORPORATION

<120> TARGETING CONSTRUCTS AND TRANSGENIC ANIMALS PRODUCED THEREWITH

<130> PXE-008.PC

<140>

<141> 1999-12-16

<150> 60/152,522

<151> 1999-09-03

<160> 39

<170> PatentIn Ver. 2.0

<210> 1

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer PGKF

<400> 1

atcgaattct accgggtagg ggaggcgctt t 31

<210> 2

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer PGKR

<400> 2

ggctgcaggt cgaaaggccc ggagatgagg 30

<210> 3

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer NeoF

<400> 3

acctgcagcc aatatgggat cggccattga ac 32

<210> 4

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer NeoR

<400> 4

ggatccgcgg ccgccccag ctggttcttt ccgcctc 37

<210> 5  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer TKF

<400> 5  
ggatcctcta gagtcgagca gtgtggtttt 30

<210> 6  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer TKR

<400> 6  
gagctcccgt agtcaggttt agtcgtccg 30

<210> 7  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer F5R51

<400> 7  
gtacatttaa atcctgcagg 20

<210> 8  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer F5R52

<400> 8  
agctcctgca ggatttaa 20

<210> 9  
<211> 77  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer F3R31

<400> 9  
ggcccgggct taattaatgc atcatatggt accgtttaaa cgcgcccgca agcttgatga 60  
cgcgcccgcc gccggcc 77

<210> 10  
<211> 77  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer F3R32

<400> 10

gatcggccgg ccggcgccg gtcgacaagc ttgcgggccgc gtttaaaccg taccatatga 60  
tgcattaatt aagcccg 77

<210> 11

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer VN1R

<400> 11

ctgtatttaa atctgcccac cctattcagg acagtagtc 39

<210> 12

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer VN1F

<400> 12

ccaatgcatc aaccagcca ggaggagtgc g 31

<210> 13

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer VN2R

<400> 13

aacgcgtcga cttcggagat gtttcgggga taaccagg 38

<210> 14

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer VN2F

<400> 14

ttggcgccgc ccatagagaa gagacaccaa aggcacgctc 40

<210> 15

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer FosB1F

<400> 15

ctgtatttaa atcccgtttc tcaactgtgcc tgtgtc 36

<210> 16  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer FosB1R

<400> 16  
gtctcctgca ggcttcctcc tccttggtcc ttgcg 35

<210> 17  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer FosB2F

<400> 17  
aacgcgtcga cggatgggat tgacccccag ccctc 35

<210> 18  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer FosB2R

<400> 18  
ttggcgcgcc ccttgctcctcc acctctcaaa tgc 33

<210> 19  
<211> 31  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer VF1

<400> 19  
acctcactct cctgtctccc ctgattccca a 31

<210> 20  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer VR1A

<400> 20  
gctctggcgg tcacccccaa aagca 25

<210> 21  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer VF2

<400> 21  
ccctttccaa gaccggtgcc atttgagc 28

<210> 22  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer VR2

<400> 22  
actttgcccc tgccctctc tctgttcgc 29

<210> 23  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer KF1

<400> 23  
gctgcgtcca gatttgctct cagatgcg 28

<210> 24  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer KR1

<400> 24  
ttctcaggca cagactcctt ctccgtccct 30

<210> 25  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer KF2

<400> 25  
cagatggacg agaaaacagt agaggcgttg gc 32

<210> 26  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer KR2

<400> 26  
gaggactcag ggcagaaaga gagcg 25

<210> 27  
<211> 29  
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer TF3

<400> 27

agcttagcct gcaagggtgg tcctcatcg

29

<210> 28

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer TF2

<400> 28

caaatgcacc ccagagaaca gcttagcctg c

31

<210> 29

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer TR1

<400> 29

gctttcaaca actcacaact ttgcgacttc ccg

33

<210> 30

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer F51

<400> 30

cccagtgctc ctgatttagg gagagcacct gag

33

<210> 31

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer R51

<400> 31

ccagactgcc ttgggaaaag cgcctc

26

<210> 32

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer F52

<400> 32

cagtgaagt cttctctgtc cctcaatcgg ttctg

35

<210> 33  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer R52

<400> 33  
tggatgtgga atgtgtgcga ggccag 26

<210> 34  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer F31

<400> 34  
aatcaaagag gcgaactgtg tgtgagaggt cc 32

<210> 35  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer R31

<400> 35  
cggctcccca aaatgtggaa gcaagc 26

<210> 36  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer F32

<400> 36  
gaatccatct tgctccaaca cccaacatc 30

<210> 37  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: primer R32

<400> 37  
cgctcctct cccagtcctc cccttg 26

<210> 38  
<211> 4998  
<212> DNA  
<213> Mus sp.

<400> 38



tccaccacc tggtttctcac gtccccggcc ttccctagtta acttcatggt taaagaagcc 60  
 tcaccggggg aggggtgtggt gccacagaag gaagggtgct cccacaagcc cccagtgtct 120  
 ctgatttagg gagagcacct gagccagtg agagtcttct ctgtccctca atcggttctg 180  
 aaattcccca cttgccctcc ttatccaggg gacagggtctg cccaccctat tcaggacagt 240  
 agtcttaaac tcgtagccaa cagacttttt attgggctgg gagaaagaga tgaggctcct 300  
 gaagctcagc cgagtgggt ctgattccta cttctcagag gtccggcagc ccagccaata 360  
 ctgagcaatg gagcgtgggt agggaggatt cacagagtcc actcgccggg ttctaagggt 420  
 gactcggtag tatttgtctg aaagaaagaa tggaagaaagg gttatgtgag attctgcctg 480  
 atcctgtcca ctggtcccaa gaaggataaa ggctttttct cagaaggga agtgaacatc 540  
 caccaagcag ataattgtcac catctacagg ctgtgttcag caccagggg ccaagacctg 600  
 caggcaaggc ctagccaaaa ccagtctaag gagttagaaagg gggctccac ctccagagaa 660  
 gaaatagacg ctctgaatgg gctcgagggt ggcaggtaca agccagtcca tatcataatc 720  
 atagttgttg taggttccta gccactctc ctgcgtggag aacaaagaga accagattga 780  
 acgtgatgaa cgacgggagt tcgagctctg gctgcgtctg tggccacgcc ctcggcgtga 840  
 acgtatagcg tttcggcttc tacgcttaga cttctgtttt ttggcttggg cagagtggga 900  
 taaggagcca gtgacgtaga tgcggccggc catagcagcg tccactttcc ctggcacacc 960  
 atgccagttc cggctgatga attgggggtc tctggctcca tctgtaacag ggaaggggtt 1020  
 aatgcacttg gcagattctg gctttgattt ctccagcaag gttgtctgtc tatctattta 1080  
 tctatcttta tctatgtatc tatctatata tctatgtatc tatctatcta tcatctacct 1140  
 acctacttac ctatctatgt atctatctat ctatcatcta cctacctact tacctatcta 1200  
 cctattttatt tggttgtttg ttttctttga aacaggatct tagcacctac ctatggctgg 1260  
 tttgcaactc actatgaagc cataactggc ctcttaactc acaaagatcc acttgctgt 1320  
 gtctctgagt gctgggatta aaagcatgtg ccactacacc cagctccagt aggaccttta 1380  
 gaacacattt gctatgcctt gcctaagaca cacaactcag tccccaggcc ccagcctccc 1440  
 tgtctagagc tttttcccat cctctctcca ctgtatccct tgaatctctg ccccatccga 1500  
 aacccctcag cgcgcagccc ctcttctgct tgtgttaggc aaagtccaag gtatgggatc 1560  
 caaatagagc caagcctcat ccccaaaaag tcaacagaag caaagtctag ccagagcaaa 1620  
 cagctcttga tcgatgggtg cacagttcca gggccctccc ctggaagccc ccactatcac 1680  
 agcccagttt ccagagaaaag aagccagcct tgctctccct ccataccaga ggatctgccc 1740  
 cagaagagga gttcgaaaaat gttctcccag ctgtcccgct gaagcaaggc aaagtgtca 1800  
 aacacggctg acagagagct gccttcgcac tccctctggc tgggttgctg ctgaaattcg 1860  
 tactcccagt actgcttccc tgaggagcag aacagctggc atcaggagag atctgaccaa 1920  
 ggcagagagg aatcatggaa tagaacaggg actccaccac ctgccccctt ctctccacc 1980  
 ctgagtaccc ttgaagaagt agaccctttc cgggccactg taacgggtggg caggaagggc 2040  
 gaacgctgca tcaacattgt ctggtatgcc actgaagcct tcggagatgt ttcggggata 2100  
 accaggggtc aggaccccat cctcaaagcg ccagtactga ctaccctgaa agacagagat 2160  
 cagaaggggtg aggacatacc gctggccaca gaagcagtc tatatcctaa actggctgtc 2220  
 acctgctcct ggagtccctg actgctttgt cttcacagct cccagcacg tccatggcac 2280  
 cctttacctt gcctcagact taggtctggg accttgaaca agtaggtctt cccctgacag 2340  
 ttgatgcgag tgaaggcagc atcgatgggg ccctcaatgc cccagacatc ttggataagt 2400  
 ttgggttacc caggcctcac tgccgtctca tctagctcat agcagtactg ccctagaaca 2460  
 ggggaaactg tgtgagaagc agatgagcct aaggcagatc cgaccgccac cagacctgtc 2520  
 catagagtca cctcggaagg caaagaggga cccattcttg agatccgtga aggcgtcaaa 2580  
 gggctttcca ctgcacagtt cttcctctgg aaactcaggg gtcccttgat cagtgggtgtc 2640  
 gggccttagg atctcctcct gttgctccac tttaggcgct ggggtgcttg gctgttctc 2700  
 aggatctagg aaggctgtcg gctttagagt gccgtccgtc cgaggattta ggtcaccggg 2760  
 tggagagggt ttctcgggtt gcacaccggt gttgggtattg ttcttgggt cctccacgta 2820  
 gtcatagctc caataatcat cctctggcat agtgaacacg tcccccgcg ttactgcagg 2880  
 cagaacgggg agcagtgagt tgcaggtgtg ggaggagacc ccaggccac ccaccagggc 2940  
 tctgaactca ccttggggct tgcaactgct catgtagtgc gcacagcagc tctgatagta 3000  
 agtgcaaaag tcgtcacact gacacttctt gctggccatg aaacctgag tgcagcggcc 3060  
 cttgcatgac tctatgggag ggaatatcag gtttacagcc caatctaggg cacctgcccc 3120  
 acctgcactt ccctaggtag ccaccaatcc cctccacac cttggtcagc cagagaaacc 3180  
 catgccacca gggctagtat gaaaaagggc ctgaggggtg ccatggcagg cctctagccc 3240  
 agggccttgg caagctgggc gcggagcttc tggaatctcg ctgtcctgcc tgaaaaaaga 3300  
 agcagactga agaagagttc ctagtccctt ggggttctgc cctttatttg ctcatcctct 3360  
 ggcccagccc cattgccctc ctccaaacac agctgcagca aagggtcaca ttccagaac 3420  
 cccagcccca ggagagctgg gaaacagaaa accctcgcca agaccaaagt cagtagggtc 3480  
 acgggcagga gggataacac gcttagctta gctggggagg tggaagaaag catgtgtgt 3540  
 caccctctga gccagtcccg ttaatctccc tgagccttac tttttataaa gtgggacct 3600  
 ggtgccttgc ctcatcaggt gttgagagat tccgtgagct agaacagaca aaacgttctg 3660  
 tgccctggagt agcttccaac tcattcccat aagccgttat cgatttactg tttgatcagg 3720

```

ctaggtgctt gtcccatcct acccccgcgt tcgaatctgg atttttgggg caagaagggg 3780
ggttggggga gagctggcaa gcactttggg ggaggttttc ttttctctc ataaaagaac 3840
aaagcttcat ttctggcctc tccttgttct ctctaagctg ggtgttacag cataggaagt 3900
agtgggtcag agtctattct tctttcttta ttttttttag atttatttat tttatgtttt 3960
gtgtataagt gtctgtcac atgtgcatt gtgcaccaca tgcattgtct gtgtctatgg 4020
aggtcagaag agggctttga ataccctgga actggagttt tgaacagtta tgagctgccg 4080
tgtggatgct gagaatcaaa cccaggtcct ctgtaagaac aagtactctt aaaggctgag 4140
ccatctttcc agtcccagag cccattcctg aggccttcac taatccattg atcctcgggg 4200
gaccaccctg gccacacctt caatgacctc atttatttta aaaaaaaaaat ggactcattg 4260
ggcatacttt ctagactcac atactaagtg ggatttctct ataaagaagt gctcactggg 4320
gtagagtgcc aggttttggg ccaaattcca agcactggca cacttctgaa gcccctccgt 4380
tttctgttct gtaatcacag gcgagcgtgc ctttgggtgc tcttctctat ggaccgcagt 4440
agtctcagcg gcaaaatgaa aactaaatt ttactcccta cagacgcgtg aagcctaagt 4500
ggaaaccggc attaaagggc tttaagaatc tcaactgcca ttctttaacc atccggaggg 4560
gacgtggata catgtagcca gcttgtcttc acattttggg gagccgagcg agcggtagga 4620
aatggaagac agctctttac agccctttct acagcatctt gcacaccacc aaggggagac 4680
tggggagagg aggcggagcc aggtgtgggc gtggctggag acctggggta ggcttgccgc 4740
tgcgtcgggg gcggagcccc tgaacacctag aggcggggcg tcaaatcctt gactctgctg 4800
ctcagaggcg tggttgctgt tgagcatctt agctccgctg tgcttagatt ggagcagcgc 4860
tttgttccgg gcaccggcgt ctctaccctc ccgcgtctgg tccatgcttc tctctccctt 4920
catgcccttc ctaagtcgct gagtcccgga gctgcctccc tcttctgct tctacacttg 4980
tagcccagca cctttacc

```

&lt;210&gt; 39

&lt;211&gt; 11176

&lt;212&gt; DNA

&lt;213&gt; Mus sp.

&lt;400&gt; 39

```

gcagctgggc aaacgttggc gatgccgggtg caaagtatat acccggtggt tagcagaagc 60
tgagaacttt tagccgaaag ccggtccct aagccgaagc taggcaagta ggggaagaaa 120
aagaaaaaaa aaattccaga gaagcttcca gagcctctc ctcttccctc ttccttcaaa 180
aggactgcaa gtccgcagtc accctccacc cagcaagagt tagggcctcg aaccccggtc 240
acgtgcctc cgctcctgc cgaacgtaac gggggaccgc tgcgtaaagc gtgacgcgt 300
ggaatcctcc gtctgacgcg gggcacgcac aggcgcgagc cctccgccc gccccgccc 360
tgacgtccgg gcacgttcta ttttgaacg ccgagggcac gttgctaagg gagggggcag 420
cgtggctttg tgattggctg tcgcgcgcag ctttagccaa tcagcgttcc cttcctattt 480
gtagagcgta gctcccttcc ttgctttttg tggttcttcc cgtgctgggg gtctccaaga 540
ggagagctag gattcttgtc gcgatcgga ctcgttgtca ccccatggtc tgcgaggact 600
tgtgtggacc tggctctgtt tcataagcta gaggttttg gctgagtgtt agcgcctcta 660
agggggaact gaaggcctca tcttctcag gcacacatat acgtgctcct gagctctaga 720
cactcagctc tcccgaggtg ttcaaacact agatgagcta gcctacggag aggcagccag 780
gtggtctcta aaaggtctgc cttcccttag ttcccaggct ctgattggcc agggattcag 840
ccctccctc gccacgccc ctagagtagt taagcctcta ggattccact tgcgggaagg 900
gggggggggg gggcgtgatg gacgcttctt ggggacgcag atcctatgtc accccatccc 960
ctgcaagaca gtctgagaga ttctcgctgt cacttttctc tgcctatcag ttcactgaaa 1020
cctgtcagtc tactgggaa gagacagaca ctcggaagg atgctctcaa ctcttaggcc 1080
ggtcccccac caccgttga actgggatct ccgcctgcgg gagccctcat gcagtggggg 1140
gtgtgtttgt gtgtgagtgg agaggaaggc ttggctaagg cctctccctc tccctccctc 1200
tgtgtggggg gttggggggg tttggctgta tgtgtgtgtg aatgtctgtg gctccatccc 1260
gggagtttgt caccaggttc tgtccagcct cctctccac ccaccccccc acacctaaga 1320
gtcaccaacc cggggtgtga ttcaccaccc gctggaaccg tgcaaccttt ccccgaggaa 1380
gaaggaggag gtagaaggca gttgaacaga atctctcatt aaccactgcg tcacgggtga 1440
gtggaagggt ggggtgtgtg gctttttgcc tgtgacacac acatccacac ccgctcacc 1500
tgtgtcact cacagggtcg gtctctctta tctctcttgg gcgtgtgtgt gtcggtggct 1560
ttgtttgtgt gtctacgcct gtgtgtgtat gtctacccc gtaggagtgc gccgtctcgc 1620
gggaaatgcc cggctccttc gtgccaacgg tcaccgcaat cacaaccagc caggatcttc 1680
agtggctcgt gcaaccaccc ctcatctctt ccatggccca gtcccagggg cagccactgg 1740
cctcccagcc tccagctgtt gacccttatg acatgccagg aaccagctac tcaaccccag 1800
gcctgagtgc ctacagcact ggcggggcaa gcggaagtgg tgggccttca accagcaca 1860
ccaccagtgg acctgtgtct gcccgctccag ccagagccag gcctagaaga ccccgagaag 1920
agacagtaag tatgaggcct caggagtgtg gatggaggag cctagctagg gatgtgggct 1980

```

cagtttgtac agtgccttgc tgccatgcat gaagatccct agcacagcat aagccaggag 2040  
 tggttatgca gacctgtaac cccagctctc agaaggtgga ggcaggagga gcaggagttc 2100  
 gaggccagcc tgtgctactt atggagtcca gcctgcactg caagagatca ttattttcaa 2160  
 aagttggcct tggggggagg tgggtgaggg aagttaagaga aagtgcagct aattttgtca 2220  
 cttaatatgt ggaggttctt ctgaggcctc aagtctgaag gaactttacc attctggcca 2280  
 gtgaggagta ggggttatta tttggggttc aggaggaagg aagttttctt agggctgata 2340  
 gaggtagccc cagatctcat ggtccttate tctgactcag cttaccccag aagaagaaga 2400  
 aaagcgaagg gttcgagag agcggaacaa gctggctgca gctaagtgc ggaaccgtcg 2460  
 gaggagctg acagatcgac ttcaggcggg aaggaggagt ctgggggtgt cttgaggccg 2520  
 tgctgggagc actctgcctt gttcttcccc cgtttctcac tgtgcctgtg tcctaaacga 2580  
 ggaaaccccc tcttagggaa cagggtcag tataggctga tggagtggct ccatatgcat 2640  
 gctcagaccc atgcccactt actttcgact gttccccact tccctgaat atgtcccccac 2700  
 atgtcacctt cctggcttct tctcagccta aggagacaag ctaggaggag taattctctc 2760  
 accttctttt cttactaaa taataatcca ttttgccctc ctgcctccat ttttttttcc 2820  
 tgagctgggg atctacctgt cgtagtctag ccctcctccc ccaacttgat agcctcaagt 2880  
 ttcagccctt ggctgagatg ccatcatcct gactggctct ggctggaaac tattttgtgc 2940  
 taagtcaatt ccttgtctgc tacttcagct atctacagty ctgccgaact tgagctggtg 3000  
 gcgcccacca agcccacttc tttctctctt ttttacctca gtgcaacccc ccacacacaa 3060  
 aacttcagtc ctgccccttg aaaccagggt gcgtctctga cccccgtcg ggaggctgaa 3120  
 ggagatgggt aacagaacct cattaaaaac aacacataag cattacctac tgactcaaca 3180  
 aactgtagtg tttttctttt ttctctctaa aaaattattt cgtttgttta tttattattt 3240  
 gcttatgttt gagtgagtgc tgggtcacca cagcacacat acgaggtcag agggaaattt 3300  
 tcatagtttg ttctctcctt ccgtgttgtg ggtgcttgct ggcaatctcc ttcactcagt 3360  
 gagctacaat gccccttctt gccctttaag gcagagtact ccttagtaca gggggacctt 3420  
 ttctctcgcc tctcaaagt gagattacaa atgttcacca tcacaccagg cttggagttc 3480  
 ttgcctatca gtgacgtcca ctctgccta gcttcttccc aacctcttt tagtctgatg 3540  
 gggaaaccga ggcacgagta gcatggctca ccaggatttc ctcttagggg acggctccct 3600  
 cagttgggag ggagctgtcc agccccctgg atcagcagca agaattgatg agtgtgggg 3660  
 tgggctgggtg aagctactct gtgtggctgc tgaccagcaa ttctcctttc tctgtctcct 3720  
 atgacctggc cctgctggga tccattagga aactgatcag cttgaagagg aaaaggcaga 3780  
 gctggagtcg gagatcgccg agctgcaaaa agagaaggaa cgcctggagt ttgtcctggt 3840  
 ggcccacaaa cgggctgca agatccccta cgaagagggg ccggggccag gcccgctggc 3900  
 cgaggtgaga gatttgccag ggtcaacatc cgctaaggaa gacggcttcg gctggctgct 3960  
 gccgccccct ccaccaccgc ccctgccctt ccagagcagc cgagacgcac cccccaacct 4020  
 gacggcttct ctctttacac acagtgaagt tcaagtctc ggcgacctt tccccgtgt 4080  
 tagcccttcg tacacttctt cgtttgtcct cacctgccc gaggtctccg cgttcgccc 4140  
 cgcccaacgc accagcgga gcgagcagcc gtccgacccg ctgaactcgc cctcccttct 4200  
 tgctctgtaa actcttttaga caaacaaccc gcaaggaaca aggaggagga 4260  
 agatgaggag gagaggggag gaagcagtc ggggggtgtgt gtgtggacct tttgactctt 4320  
 ctgtctgacc acctgccgc tctgccatcg gacatgacgg aaggacctcc tttgtgtttt 4380  
 gtgtctgtc tctggttttc tgtgcccgg cgagaccgga gagctggtga ctttggggac 4440  
 aggggggtgg gcggggatga acacccctcc tgcatactt tgtcctgtta cttcaaccca 4500  
 acttctgggg atagatggct gactgggtgg gtagggtggg gtgcaacgcc cacctttggc 4560  
 gtcttacgtg aggtggagg ggaaagagt ctgagtgtgg ggtgcagggt ggggttaggt 4620  
 cgagctggca tgcacctcca gagagacca acgaggaaat gacagcaccg tctgtcctt 4680  
 cttttcccc acccaccat ccacctcaa ggggtgcagg tgaccaagat agctctgttt 4740  
 tgctccctcg ggccttagct gattaactta acatttcaa gaggttaca cctcctcctg 4800  
 gacgaattga gccccgact gaggggaagtc gatgccccct ttgggagtct gctaaccca 4860  
 cttccgctg attccaaaat gtgaaccctt atctgactgc tcagtcttcc cctcctggga 4920  
 aaactggctc aggttgatt tttttcctcg tctgctacag agccccctcc caactcaggc 4980  
 ccgtccccc cctgtgcag tattatgcta tgtccctctc accctcacc ccacccagg 5040  
 cgcccttggc cgtcctcgtt gggccttact ggttttgggc agcagggggc gctgcgacgc 5100  
 ccatcttgct ggagcgttt atactgtgaa tgagtggctg gattgctggg cgcgccggat 5160  
 gggattgacc cccagccctc caaaactttt cctgggcctc ccttcttcc acttgcttcc 5220  
 tccctccctc tgacaggag ttagactcga aaggatgacc acgacgcac ccggtggcct 5280  
 tcttgctcag gccccagact tttctcttct aagtccttcg ccttccccag cctaggacgc 5340  
 caacttctcc ccacccctgg agccccgcat cctctcacag aggtcgaggc aattttcaga 5400  
 gaagttttca gggctgaggc tttggctccc ctatcctcga tatttgaatc cccaaatagt 5460  
 ttttgacta gcatacttaa gagggggctg agttccact atccactcc atccaattcc 5520  
 ttcagtcaca aagacgagtt ctgtcccttc cctccagctt tcacctcgtg agaatccac 5580  
 gagtgcagatt tctattttct aatattgggg agatgggccc taccgcccgt ccccgctgct 5640  
 gcatggaaca ttccataccc tgtcctgggc cctaggttcc aaacctaat ccaaacccca 5700

```

ccccagcta tttatccctt tcctgggtcc caaaaagcac ttatatctat tatgtataaa 5760
taaataatatt atatatgagt gtgcgtgtgt gtgcgtgtgc gtgcgtgcgt gcgtgcgtgc 5820
gagcttcctt gttttcaagt gtgctgtgga gttcaaaatc gcttctgggg atttgagtca 5880
gactttctgg ctgtcccttt ttgtcacttt tttgttggtg tctcggctcc tctggctggt 5940
ggagacagtc ccggcctctc cctttatcct ttctcaagtc tgtctcgctc agaccacttc 6000
caacatgtct ccactctcaa tgactctgat ctccggctctg tctgttaatt ctggatttgt 6060
cggggacatg caattttact tctgtaagta agtgtgactg ggtggtagat tttttacaat 6120
ctatatcgtt gagaattctg ggtggaaatg tctgatcagg agaagggcct gccactgccg 6180
accacaattc attgactcca tagccctcac ccaggctgta tttgtgattt ttttcatttt 6240
gtttttttgt attttgcacc tgaccccggtt ggtgctgggg cagtctatca ctgggcagct 6300
ccccccccc ccttggttct gcactgtcgc caataaaaag cttttaaaaa actgtatcct 6360
tcagggtcaaa gtgtctggtt tccctggaca tctactacat ggcttccttt cagaaaaacg 6420
gagtttgatg tctagggaa gtcttgctgg cacttagtgg gacgcctaac gaatcagaac 6480
ctacaacggg actaaaagga agtggagact tgctaggttt tcccatgttc ccaggctggg 6540
ccacctactt gaaaaataaa ggggcggaag agtctaaggt accaaatttg gtgaagggtc 6600
tgggagaatt tcatgatcgg aaaagaattt attcaccttg ggtgtgcaat gaactttcag 6660
caacagttaa gggcaagggt gtaaaagctg ggcacaactt gtaaatccta gcatttgaga 6720
ggtggaggca aggggatcaa ctggtggagt tcagtgtcat gtggatcgta gataccaagc 6780
gcaaagatct gctatgggga gagggcttgg tacaccaggg gagccagaag tttcgtgggtg 6840
agggtagtgg agggcaagtg gagagtgaga gttagcctca gggagattct acaggcaatg 6900
atgcagagtt cagacgctcc ctttgaaagc actagagagc cgcagcaggt tttgagcaga 6960
gaaggttaga gttaggtggt ctcttctagc ccatcccagg ctgaggaggga cgctgagggt 7020
ttcaagaagg atcgagaatg gaaagcagag gagaagaagg atccaagagg catggaggag 7080
gcagaacaca tttctcttct ttaatagcaa gcctggaaa gataacttgc tgcaggaggga 7140
gatgctcacc agtcgggttg tctaggggggt tcttggaaaa gagaaggcat ttgctcaagc 7200
ctcgggtccc ccattctcgc tcttctgtca gcttgtcttc cattaagtgt gtgtctcaag 7260
gccaccctgc tcaggactcc ttgtgagacg accttctatg ctcgagtcca ttaaaaaacac 7320
aattgcctgg tgccgtgtct tctccactgg ctccagtacc tcaaaagacc agggctaaag 7380
gtgtgatcac aactctatcc ccattactgc tccaacgcag agacaggact gagccggagt 7440
gaacaaatga acaaaaatga ctaataatgc atgcgtgatt aaatacataa aagagcagat 7500
gactggatga gcaaatcggt taaggagaga cagcaagatc ctagaatttt ggagactaat 7560
ttaaatccat ctttgagatg catttggctg gaaattcctg ggaggaaaaa aagtgtaaat 7620
atgaagagag aataaatgag aataggggtg gcttcagaga ggttaactgc gcgtggctg 7680
cttttgatca agaattgtga ttgcaggag caaaatggga tagatactcc cgcccgaaag 7740
gtggaattga accactctgt cgctaaacag ctacagggtt gaagcctgca cccagacca 7800
ctgaggatca tccgggcgaa aggagctatt ttcagttagt tatataaagg cgagatacta 7860
ctacttttta cacttatggt cattatttgt ggtatacagt agataattaa tttcaatggt 7920
ttcgaacatt ttttttact ttttctgtg aacatgtgtt tcctcagtaa agtgttccgt 7980
gaatgactct actaactaaa agttaagtag cttcatttgc atagcgcctt gcattttggg 8040
aagcagcgcc taaagtgcct gtctccctaa ctaaaagcag aattttttgc aaagtgaana 8100
gtcagtttta tttttgtttg tttgtttgct tgtttgtttt taatggaaaa acttctcacg 8160
cggcccatte gtagcagaat tcgagatttt ctgcaagcga gaagcaagac tttcgtaggg 8220
tctgacggca cgcggccgca gagcgacacc tgccgttgct ttatagaact gcaagtatgt 8280
agggaaatcta ctgagtcctt aggtgatgga gttgacaacc aactccctt gagttttagac 8340
gctaaaaacc atcccttttt atatttatgt gattagccca gggaaactaa ggctcagaca 8400
tggataatac cacagccgag ttctttagc ccaactccct aggggaaatg aaacctacag 8460
ttgtggtttt aatatgcttg gccagggggc agtggcccta ttggcaggag tggccttatt 8520
agcggagggtg taccttggtg gagaagtgtg tcaactggag gcgagggttt gaggtacgta 8580
tgctcaagtc tggccagtgt gatcctggct gtctgcagaa cgtgggtctc tctggtgctg 8640
cttcggatca aggtgtagaa ctctcagctc cttctccagc accatgtctg cctgcttaat 8700
gctttgcttc tttccatgac gataatgaac tgtgctctg aaactgtaag tcagcccccc 8760
agttacatgt tttcttttat aagagttgca tatatatatg tatgtatata tgtatgtata 8820
tatgtatgta tatatatata tatatataaa cagggtctca ctctttagct ctggctggcc 8880
tgaaattcac tatgtagccc aggattgcct gaactttgaa gcaatcttcc tgcctcagcc 8940
tcccaatggt attacaggca tgagtcacaa caagccattt aaatcttatg atgacttata 9000
agaagacaga aaatcagagt tcctttacct agttcacaga tccctacaat ctaacctcgt 9060
tcgctccata aacagcccta cccaccctc ctggaactgc tttgagggaat gctgcaggct 9120
ctcacaggca cactcctcct tgggttaatct cttcagcctg gttgccttcc ccccccattg 9180
ccatgtggcc caaagcctct catcctgttc tcaaatacca ctgctagta aggtcccccg 9240
acctgaccgg gtttaaatat tagaaaaggg tcaatttctc cctgccacag acaaccaaac 9300
caccatattg ttgtcactta ctacctgact atgaagggtt atagatgtct tcacaacctt 9360
tctctgagcc tcagtttccc cacctgcata atgcatctga gacacagaat tccctagagc 9420

```

tgtggttctc ctcattccta gtgctgggac cctttaatac atttcctcat gttgtggtga 9480  
ccccaccacc accataaaat tatttccatt gatacttcat aactgtaatt ttttctattg 9540  
ttatgaatag taatgtaagc atttggtgtt cccagtgate ttagatgacc ctgtggaaga 9600  
gtcattccac cccaaagggg tccccaccac aagttaagaa ttcttgccat agaggaatca 9660  
cagggaacct ggattaacac ttgggtcgac ttttgggctg ccttctggga ggcgctagag 9720  
ctaattgacag ctacatcaat ttctgaaatt ttgtgtgtgt gtgtgtgtgt gtgtgtgtgt 9780  
gtgtgtgtgt gccctgagtc ggggtgctgag ataggccagt ggcttttagt ttcctggacc 9840  
cattactcac cagaactctc ccctcacctg attctttgat gtgaacacta tgtcttcata 9900  
gtggcggtgg caatagcagc aacagtgaac taaattttaa aagtagaact cagctggaga 9960  
tacaaatatt gcagttttga agttgggggt gattgtctaa taacttaata acataaccca 10020  
gaagagaggc cccttgggtc tgcaaaactt atatgcctca gtacagggga acgccagggc 10080  
caagaagtgg gagtgggtgg gtaggggagc aggggtgggg gagggatatag gggactttcc 10140  
ggatagcatt tgaatgtaa atgaagaaa tatctaataa aaatttgaaa aaaaatgta 10200  
ccccagtttg gcctggatct cactacctca accagactgg catgtgactc tgctgagatc 10260  
tgcctacttc tgccctcctg gtgcagaaga caatttttg aagttagtcc tcttcttcca 10320  
tcttgtggat tccagggtt gaactcgggt catcaggctt ggctgcaagt gacttactta 10380  
gggtgtctcc agaccctctc ggtttgatta gtttagatgt gcacttcatg cctgactttc 10440  
gcactatgta gatagagcaa tgtctataac atctcctaca atgatatgta tatcaagagc 10500  
caagtatga gatggctcag tgggtaagag cacagactgc tcttccaaag gtcccagatt 10560  
caaatcccag caatcacata gtggcttcca tccctctta tggaatgtct gaagactgct 10620  
acagtgtact tacatataat aaataaataa atcttaaaaa aaaaaaaccg agccgggctg 10680  
gggtggcgac gcctttaatc ccagcacttg ggaggcagag gcaggcggtt tcctgagttc 10740  
gacgccagcc tgggtctacag agtgagttcc acgacagcca gaactacaca gagaaaccct 10800  
gtctcgaaaa aaaaaagaga gagaggggaag tgagagcgca ataacttta catttctgtg 10860  
gttgtctttg ctgtagtcta ttctgataag caatgctggc ttgctccaa ggtaggaagt 10920  
aacatttctt tataaaagggt atttgctctg ctttattttt ctgttttatt tatgggtgtg 10980  
aggatggaac ccaggacctc tggcaagcaa ggctagctgt ttaccactga gccatactcc 11040  
agccttgac tgggggattc taggcaaggg ttctaaccact gagccacact cccaccccc 11100  
atccctctct ggaagattct aggcagttcc atacctagcc tttgatcttt taagacggtc 11160  
ttactagagc tcagtt 11176

**A. CLASSIFICATION OF SUBJECT MATTER**  
 IPC 7 C12N15/90 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 30715 A (ISACOFF EHUD Y ;SIEGAL MICAH S (US); UNIV CALIFORNIA (US); CALIFOR) 16 July 1998 (1998-07-16) page 29, line 5 -page 30, line 10; claims ---	1-8,10, 22-26, 32,34,35
X	WO 98 28971 A (LINK CHRISTOPHER ;UNIV TECHNOLOGY CORP (US)) 9 July 1998 (1998-07-09) page 22, line 3 - line 6; figure 1	1,2,8, 10,12, 14-17
Y	page 26, line 24 - line 28; claims 11-19; figure 1; example 3 page 18, line 9 - line 13 --- -/--	1,2,7, 10-12, 14,17-20

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"Z" document member of the same patent family

Date of the actual completion of the international search

26 June 2000

Date of mailing of the international search report

03/07/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3018

Authorized officer

Chambonnet, F

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CONTAG CH, ET AL.: "Visualizing gene expression in living mammals using a bioluminescent" PHOTOCHEMISTRY AND PHOTOBIOLOGY, (1997) VOL. 66, NO. 4, PP. 523-531. , XP000920570 the whole document ---	1,2,7, 10-12, 14,17-20
A	CONTAG PR, OLOMU IN, STEVENSON DK, CONTAG CH. : "Bioluminescent indicators in living mammals." NAT MED. , vol. 4, no. 2, February 1998 (1998-02), pages 245-247, XP000914588 the whole document ---	1
A	WO 98 36081 A (ANGELIS DINO A DE ;SLOAN KETTERING INST CANCER (US); MIESENBOCK GE) 20 August 1998 (1998-08-20) the whole document ---	1
A	WO 98 23633 A (CORNELL RES FOUNDATION INC) 4 June 1998 (1998-06-04) the whole document -----	1

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9830715 A	16-07-1998	AU 5090498 A	03-08-1998
WO 9828971 A	09-07-1998	AU 5619298 A	31-07-1998
		EP 0948612 A	13-10-1999
WO 9836081 A	20-08-1998	EP 0981633 A	01-03-2000
WO 9823633 A	04-06-1998	AU 7410398 A	22-06-1998